

OM of: US-09-303-518D-651 to: PIR-71.* out_format: pfs
Date: Jun 30, 2002 8:00 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frim+np2.model -DEV=xlh
-O=Cgnt2.1/USPTO.spool/US09303518/runat.28062002.142713.4317/app_query.fasta.1.23501
-DB=PIR-71 -OEMT=fastan -SUFFIX=pr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPEXT=7.000 -START=1 -MATRIX=blissum62 -TRANS=humad40.cdi
-DELNET=7.000 -START=1 -MATRIX=blissum62 -TRANS=humad40.cdi
-LIST=100 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0
-ALIGN=45 -MODE=LOCAL -OUEMT=pf -NORM=ext -HEADSIZ=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09303518.@CGN1.1.1092
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPEXY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518D-651
Query length: 4350
Database: PIR-71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 504.980000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
PIR2:CG4057	+ 2471.50	2805.42	5.0e-149	709	Iga-specific metalloendopeptidase
PIR2:BA1859	+ 1554.50	1753.78	8.7e-91	1545	Iga-specific metalloendopeptidase
PIR2:SA61314	+ 1551.50	1750.27	1.3e-90	1541	Iga-specific metalloendopeptidase
PIR2:SA7023	+ 1536.50	1733.30	1.2e-89	1541	Iga-specific metalloendopeptidase
PIR2:BA1018	+ 1532.00	1717.45	9.9e-89	1431	Iga-specific metalloendopeptidase
PIR2:BA4106	+ 1519.00	1712.51	1.6e-88	1694	Iga-specific metalloendopeptidase
PIR2:BA1859	+ 1517.00	1710.18	2.1e-88	1532	Iga-specific metalloendopeptidase
PIR2:SA6039	+ 1508.00	1700.88	7.7e-88	1532	Iga-specific metalloendopeptidase
PIR2:CA1859	+ 1475.00	1661.93	1.3e-81	1773	Iga-specific metalloendopeptidase
PIR2:BA1937	+ 1470.00	1657.93	1.3e-81	1773	Iga-specific metalloendopeptidase
PIR2:CG1169	+ 1404.00	1580.88	3.2e-81	1815	Iga-specific metalloendopeptidase
PIR2:SA4632	+ 1083.50	1219.41	5.7e-61	1377	Iga-specific metalloendopeptidase
PIR2:SA7664	+ 1003.00	1126.64	8.4e-56	1366	Iga-specific metalloendopeptidase
PIR2:SA61339	+ 870.00	984.04	2.0e-47	508	Iga-specific metalloendopeptidase
PIR2:SA61330	+ 862.00	974.93	6.4e-47	508	Iga-specific metalloendopeptidase
PIR2:SA61338	+ 855.00	966.95	1.8e-46	508	Iga-specific metalloendopeptidase
PIR2:SA61331	+ 847.50	968.43	1.9e-46	507	Iga-specific metalloendopeptidase
PIR2:SA61335	+ 844.00	954.42	8.8e-46	508	Iga-specific metalloendopeptidase
PIR2:SA61337	+ 843.00	953.50	1.0e-45	496	Iga-specific metalloendopeptidase
PIR2:SA61337	+ 842.00	952.36	1.2e-45	496	Iga-specific metalloendopeptidase
PIR2:SA61337	+ 840.00	950.08	1.6e-45	496	Iga-specific metalloendopeptidase
PIR2:SA61335	+ 834.00	943.25	3.8e-45	496	Iga-specific metalloendopeptidase
PIR2:SA61332	+ 815.00	921.60	6.1e-44	496	Iga-specific metalloendopeptidase
PIR2:SA61332	+ 760.00	859.59	1.9e-40	462	Iga-specific metalloendopeptidase
PIR2:SA61333	+ 687.00	776.26	8.0e-36	471	Iga-specific metalloendopeptidase
PIR2:SA61337	+ 631.50	703.88	3.1e-32	1300	Iga-specific metalloendopeptidase
PIR2:BA15004	+ 444.50	501.85	1.9e-30	384	Iga-specific metalloendopeptidase
PIR2:CG1068	+ 372.50	407.14	8.7e-16	1569	Iga-specific metalloendopeptidase
PIR2:CG1068	+ 355.00	387.20	1.1e-14	1571	Iga-specific metalloendopeptidase
PIR2:CG1068	+ 354.00	386.31	1.3e-14	1571	Iga-specific metalloendopeptidase
PIR2:CG1068	+ 351.50	353.58	1.4e-12	949	Iga-specific metalloendopeptidase
PIR2:BA95206	+ 298.50	312.81	5.1e-11	4776	Iga-specific metalloendopeptidase
PIR2:BA95206	+ 285.50	309.54	2.8e-10	1349	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 276.50	298.54	1.0e-09	1356	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 263.00	285.02	6.9e-09	1356	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 260.00	272.98	1.3e-08	1341	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 257.00	272.98	1.3e-08	1341	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 255.00	263.68	2.9e-08	4558	Iga-specific metalloendopeptidase

PIR2:BA3855	+ 254.00	272.70	2.9e-08	1477	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 252.00	266.54	4.1e-08	1271	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 250.50	269.01	4.7e-08	1230	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 250.00	269.64	5.0e-08	1450	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 250.00	269.64	5.0e-08	1250	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 247.50	262.47	7.8e-08	1200	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 246.00	255.29	1.1e-07	3705	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 245.50	266.73	9.3e-08	978	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 244.50	267.00	1.0e-07	836	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 241.50	254.67	1.9e-07	1249	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 240.50	257.26	2.1e-07	1487	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 239.50	261.31	2.2e-07	836	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 239.50	261.31	2.2e-07	836	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 236.00	253.17	3.9e-07	1325	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 234.50	245.15	5.4e-07	2554	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 234.00	245.10	5.4e-07	928	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 234.00	247.09	5.6e-07	2021	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 233.50	247.03	5.9e-07	1910	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 233.50	245.20	6.1e-07	1240	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 231.50	248.39	7.5e-07	1275	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 230.50	241.55	9.5e-07	2299	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 230.00	248.41	9.0e-07	1052	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 229.50	240.51	1.0e-06	2529	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 229.50	246.38	1.0e-06	1238	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 227.00	246.18	1.4e-06	922	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 227.00	238.22	1.6e-06	2232	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 226.50	245.73	1.5e-06	910	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 224.50	241.82	2.0e-06	1091	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 224.00	242.87	2.1e-06	911	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 221.50	238.93	3.1e-06	1029	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 221.00	238.01	3.4e-06	1070	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 220.50	244.17	3.2e-06	507	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 220.50	232.80	3.9e-06	1791	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 220.50	229.10	4.1e-06	2273	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 219.50	223.77	4.3e-06	2535	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 219.00	245.30	3.8e-06	370	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 219.00	230.18	6.0e-06	6713	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 217.00	229.12	6.0e-06	1649	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 217.00	229.96	6.4e-06	1577	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 215.50	222.14	8.6e-06	2514	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 215.00	222.18	9.5e-06	2902	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 214.50	228.93	9.0e-06	1286	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 214.50	228.35	9.1e-06	1374	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 214.00	226.69	9.3e-06	1651	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 214.00	229.74	9.4e-06	1106	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 213.50	221.63	1.2e-05	2551	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 212.00	228.55	1.2e-05	880	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 211.00	217.89	1.7e-05	2817	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 210.50	222.23	1.7e-05	1635	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 209.00	222.13	1.9e-05	980	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 206.50	222.51	2.8e-05	955	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 206.00	218.44	3.1e-05	1343	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 206.00	218.71	3.2e-05	1367	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 205.50	218.30	3.4e-05	1343	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 205.00	229.35	3.0e-05	370	Iga-specific metalloendopeptidase
PIR2:BA3855	+ 205.00	218.71	3.6e-05	1205	Iga-specific metalloendopeptidase

seq_name: PIR2:CG4057

seq_documentation_block:

Iga-specific metalloendopeptidase type 1 homolog - Haemophilus influenzae (strain Rd)

CSpecies: Haemophilus influenzae

CDate: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997

CAccession: C64057

R.Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirsness, E.F., Kerlavage, D.M., Brandon, R.C., Scott, J., Shirley, B., Liu, L.I., Glodex, A., Kelley, J.M., Weidman, Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. A:Accession: C64057 MUID:95350630

543 alaShnHsAsnThrThrGlnValAlaAsnIleThrIleThrGlyAsnGlu 559

1693 AGATTATCACACACCGAGTGT...AGAAATATCAATAGACTTAATTACAG 1739

560 SerIleThrAlaProSerAsnLysAsnIleAsnLysLeuAspTyrSe 576

1740 CAAGAAATTCGCTACAAACGGTGTGGCGAGCAAAAGTACGACCAAA 1789

576 rLysGluIleAlaTyrAsnGlyTyrPheGlyGluThrAspLysAsnLysH 593

1790 CGAACGGCGGCGCTCAACCTGTGTTTACAGCCCGCCGCAAGACCGCAC 1839

593 lAsnGlyArgLeuAsnLeuIleTyrLysProThrThrGluAspArgThr 609

1840 CNGCTGCTTCCGCGCGCAACAATTTAAACGGCAACATCACGCAACAAA 1889

610 LeuIleuLeuSerGlyGlyThrAsnLeuLysGlyAspIleThrGlnThrLys 626

1890 CGGCAACATGTTTTCAGCGGCGAGCGACACCGCGCGCTACATCATCT 1939

626 sGlyLysLeuPhePheSerGlyArgProThrProHisAlaTyrAsnHsL 643

1940 TAGGAAGCGGGGTGTCAAAATATGAAGGTATCCCAAGAGGAATTCGTG 1989

643 euAspLysArgTyrPheSerGluMetGluGlyIleProGlnGlyGluIleVal 659

1990 TGGGACCAACGACTGATCNACCGCAGTTTAAACGGCAAAATTTCCATAT 2039

660 TrpAspTyrAspTyrPrlLeuAsnArgThrPheLysAlaGluAsnPheGlnI 676

2040 TCAGGGCGGGCGAGGGGTGATTTCCCGCAATGTCCCAAGGAGGAAGCG 2089

676 eLysGlyGlySerAlaValAlaValSerArgAsnValSerSerIleGluGlyA 693

2090 ATTGGCATTTGGACATCACCGCCCAAGCAGTTTGTGTGTCGACCGGAT 2139

693 snTyrThrValSerAsnAsnAlaAsnAlaThrPheGlyValValProAsn 709

seq_name: pir2:B41859

```
852 CCTTATTCGGCAGGAAACGGTTCCAGCTGATACGCAAGATGGT 901
| ||| :||| ||| :||| :||| :||| :||| :||| :||| :|||
312 rAspYrTrpAlaGlyTyrAsn.....LysLysSerTrpG 324
902 TCTACAGTACATTTACAGAGCGGATACATACCGCTCTTTTGAACG 951
324 lnglUtrpAsnIleTyrLysProGluPheAlaGluLysIleTyrGluGln 340
952 CGCAGTAACGSA.....CATTTTCCTTACATC 980
||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
341 TyrSerAlaGlySerLeuIleGlySerIleThrAspTyrSerTrpSer 357
981 CACAACAACGGGTACGGGTACGGTACAGAAACCAACGAAAGTNTCCA 1030
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
357 rAsnGly..... 359
1031 ATCCAAAGCTTAATACAGACAGCTCCAGCTTTGACGAATCTTGAAAT 1080
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
360 .....LysThrSerThrIleThrGlyGlyGlyLysSerLeuAsn 372
1081 GAATCGATTAAGAACAGATTATACGGCGGCGGCTTTAATCAGTACCG 1130
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
373 .....ValAspLeuAlaAspGly.....Lys 379
1131 TCCAAGCTTAACAACGGTGAACACCTTTCTTATTCGATTACGGCAACG 1180
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
379 sAspLysProAsnHisGlyLysSerValThrPhe.....GluGlySerG 394
1181 GCAACCTGATCTTATCAACAACATCAACAACGAGCGGCGGCTTGAT 1230
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 lYThrLeuThrLeuAsnAsnHisGlyLysValAlaGlyLysLeuPhe 410
1231 TTGGAAGGTGATTTTACGTC...TCGCTGAAAAACAACGAAGTGGCA 1277
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
411 PheGluGlyAspTyrGluValIleGlyLysThrSerAspAsnThrTrpLys 427
1278 AGGCGCGGCGCTTCATACAGTGAAGACAGTACCGTTACTTGAAGTAA 1337
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
427 sGlyValAlaGlyValSerValAlaGluGlyLysThrValThrTrpLysValH 444
1338 ACGGCGTGGCAACGACCGCTGTCAAAATCGGCAAAAGGACCGCTCAC 1377
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
444 lAsnProGluIleThrAspArgLeuAlaLysIleGlyLysGlyThrLeuIle 460
1378 GTTCAACGCAAGGGAACCAAGGCTCGATCAGCTGGCGGACGCTAC 1427
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
461 ValGluGlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyTh 477
1428 AGTATTGTTGATCAGCAGCAGCAGATTAAGGCAAAAACACCTTTA 1477
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 rValIleLeuLysGlnGlnThrAsnGlySerGly...GlnHisIleAlaPhe 493
1478 GTGAATTCGCTGNTCAGCGGAGGAGTACGGTCAACTGAATGCGCAT 1527
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
493 lAsnValGlyIleValSerCylThrSerThrLeuValIleuAsnAspAsp 509
1528 AATAGATTCAACCCGACAAACTCTATTGCGCTTTCGCGGCGAGCTTT 1577
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
510 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyLysArgGle 526
1578 GGATTTAAGCGGCAATCGCTTGGTTCACCGGATTCAAATAATACGATG 1627
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
526 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspG 543
1628 AAGGCGGATGATTCGNCATCAATATCCACAAACATCACCACGTTAC 1677
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
543 lnglYAlaArgLeuValAsnHisSerThrSerLysHisSerThrValTh 559
1678 ATTAACAGGGAATGAAGTATTACAAACGAGTGGTAAGAATATC..... 1722
||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
560 lIleThrGlyAspAsnLeuIleThrAspProAsnValSerIleTyrTrp 576
```

```
1722 ..... 1722
576 rValLysProLeuGluAspAsnProTyrAlaIleArgGlnIleLysTr 593
1722 ..... 1722
593 yrglYTrGlnLeuTyrPheAsnGluGluAsnArgThrTyrTrpAlaLeu 609
1722 ..... 1722
610 LysLysAspAlaSerIleArgSerGluPheProGlnAsnArgGlyLysLe 626
1722 ..... 1722
626 rAsnAsnSerTrpLeuTyrMetGlyThrGluLysAlaAspAlaGlnLysA 643
1723 .....AATAGACTTAATTACAGCAAGAAATTCCTACACAGGTTGG 1764
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
643 sAlaMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyr 659
1765 TTTGGCGGAGAAGATACGACCAAAACGAGCGGCGCTCAACCTGTGTTA 1814
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
660 PheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLeuAsnValThrPh 675
1815 CCAGCCCGCGCAGAGACCGCACCCGCTGCTTCGCGCGCAACAAT 1864
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
675 eLysGlyLysSerGluGlnAsnArgPheLeuLeuThrGlyThrAsnL 692
1865 TAAACGCAACATCACGCAAAACGCAAACTGTTTTCAGCGGCGACA 1914
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
692 euAsnGlyAspLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArg 708
1915 CCGACCGCGCGCTACAAATCATTTAGAACGCGGTGCTCAAAATGA 1964
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
709 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 725
1965 AGGT.....ATCCCAAGAGAAATCGTGGGCAACGACGATGACN 2008
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
725 pSerHisPheSerGluAsnAsnGluValAlaValGluAspAspTrpIleA 742
2009 ACCGCACTTTAAAGCGGAAATTTCCATATTTCAGGCGGCGGACGCTG 2058
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
742 sNArgAsnPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeu 758
2059 ATTTC...CGCAATGTCGCAAGGAGGAGCGATTCATTTGACGCA 2105
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
759 TyrSerGlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAs 775
2106 TCAGCGCCCAAGCAGTTTGTGTGCGCACCGCATCAAGCCATCAATCT 2155
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
775 nAsnAlaLysValHisIleGlyTyr...LysAlaGlyAspThrValC 790
2156 GTACAGCTTCGAGCTGACNGGCTGACAAATTTGTGTGCAANAANCATT 2205
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
790 ySValArgSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeu 806
2206 ACCGACGATTAAGTATGCTTCATTGACTAAGACGACGACGACGACGAC 2255
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
807 Ser...AspLysAlaLeuAsnSerPheAsnProThr..... 817
2256 TGTNAGCTNNCCNATNACGNTNNTTMAAANCCTCNCNGCGCTGNNCNC 2305
817 ..... 817
2306 TNAANGCAATCTAGTGCAAAATGGCGATACAGTTATACAGTACGCCAC 2355
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
818 .....AsnLeu..... 819
2356 AACGCCACCAAAAGCGCAACCTTACCTGCGGCAATGCCCAAGAAC 2405
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
820 .....ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheVal 832
2406 ATTTAATCAGCCACATTAACGGCAACNCATCGGNTTCGGGCAATGCTT 2455
```


832	lleuGIlylAlaAlasnleuphegIlythrIleGlnserArgGlyAsnSerg	849
2456	CATTAACTCTAAGCAGACACAGCCCGCAAAAGGACAGTCTGACGCTTCC	2505
849	lnValArgleuThrGluAsn.....	855
2506	GACAAACGCTGAAGCAAAACGTAAAGCATTCCGACACTCAAGGCAATGTCTC	2555
855	855
2556	CCTAGCCGATPAAGGACAGATTCATTCCATTGTGAAAAACAGCCGTTTACCGGAC	2605
855	855
2606	AAGTACAGCGGCACAGAGNACAGATTCACCTTAAAGACAGCGAATCG	2655
856SerHisTrp	858
2656	ACGCTGCCTGACAGCAGCGAATTAGCAATTAAACCTTGCACAGCCGAC	2705
859	HisleuThrGlyAsnSernsryAlaHisGlnleuAspLeuAlaAsnGlyHis	875
2706	CATTACACTCAATTCGCGCTATCGCAGATGCTCTCAGGCGCGCAACCG	2755
875	slleHisleuAsnSerAlaAspAsnSerAsn.....	885
2756	GCAGNGTGTTCAGACAGCGCCGCGGCTTCGCGCGTTCCCTATTAATCC	2805
886Asn	886
2806	GTTACACCGCCAACTTCGGTAGAATCCGGTTTCAACAGCCTGACGGTAA	2855
887	ValThr.....LysTyrAsnThrleuThrValAs	896
2856	CGCAATTTGAACNGTCAGAAACATTCGCGTTTATGTCGGAAGCTCTTCG	2905
896	n...SerleuSergGlyAsnGlySerPheTyrTyrleuThrAspSernera	912
2906	GCTACCGAAGCAGCAAAATTGAAGCTGGCGAAAGTTCCGAGAGNACTTAC	2955
912	snLysGlnGlyAspLysValValThrLysSeralThrGlyAsnSph	928
2956	ACCTTGCGGCTCAACAATTCGCGGACAGACCCGTAGAGCTTCATTA	3005
929	ThrleuGlnValAlaAspLysThrGlyLeuPro...AsnHisAsnGluLe	944
3006	GAGCGTAGTGGAGGAGGAAAGACAAACAACCGCTGCCAAACCTTAAT	3055
944	uThrleuPheAspAla...SerLysAlaGlnArgAspHisleuAsnV	959
3056	TCAACCTGCAGAAACGACACGTCGATGCCGCGCGGTGGCGTTTACCAACTC	3105
959	alSerleuValGlyAsnThrValAspLeuGlyAlaThrPlyLysTyrLeu	975
3106	ATCCGCAAGAGCGAGAGTTCGCGCTGGATTAATCCGTCAGAAAGCAAA..	3153
976	ArgAsnValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysAr	992
3153	3153
992	gaSnglnThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnA	1009
3154GAGCTTCGCGCAAACTCGGAG..	3177
1009	laAspValProSernValProSernAsnGlnGluIleAlaArgValAsp	1025
3177	3177
1026	gluAlaProValProProAlaProAlaThrProSergIuThrGln	1042
3178GAGAGGCGCAAAAGACGGCGGAAAA..GACACG	3211

1042	uThrValAlaGlnAsnSerLysGlnIleuSerLysThrValIcLysAsnG	1053
3212	CGCAAGCCTTGACGGCGCTGATTGGCGGCGCGAGATCCGCGCA...	3258
1059	IuglnAspAlaThrGluThrThrAlaGlnAsnArgIleValAlaLysGlu	1075
3259AAGACGAAAGGGTGGCCGACG	3281
1076	AlaLysSerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..	1091
3282	GGCCCGGCGGACGGCGGGGAAATGTCGGCATTCAGACGGAG...	3327
1092SerGlySerGluThrLysGluThrGlnThrGluThrL	1105
3328GAAGAAAAAAGCGGTGACGGCGAT	3354
1105	ysGluThrAlaThrValIcLysGlnGluLysAlaLysValIcLysGlu	1121
3355	AAAGACAGCGCMTTGGCGAAA.....CAGCCGCAAGCGGAACCG	3395
1122	LysThrGlnIleuValIcProLysValThrSerGlnValSerProLysGlnI	1138
3396	GCCGGATTACCAAGCGCTTCCCGCGCGCCGCGCGCGGATTTGC	3445
1138	uGlnSerGluThrValGlnProGlnAlaGluProAlaArgLysAsnAsp	1155
3446	CG.....CAACGCGAGCCCA.....	3452
1155	roThrValAsnIleLysGluProGlnSerGlnThrAsnThrAlaAsp	1171
3463CCGCAACCT.....	3471
1172	ThrGluProAlaLysGluThrSerSerAsnValIcGlnProValTh	1188
3471	3471
1188	rgLysThrThrValAsnThrGlyAsnSerValValIcLysnProGluA	1205
3472CAACC.....	3477
1205	snThrThrProAlaThrThrGlnProThrValAsnSerGluSerSerAsn	1221
3478	CAACCGCAGCC.....	3489
1222	LysProLysAsnArgHisArgArgSerValArgSerValProHisAsnVal	1238
3490G	3490
1238	IcLysProAlaThrThrThrSerSerAsnAspArgSerThrValAlaLeuGlyA	1255
3491	ACCTGATTNAGCGTTATGGCAATAGCGGTTGAGTGAATTTCCGCC...	3537
1255	spLeuThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLys	1271
3538ACCGTCAACAGCGTTTCCCGCGTACAGACGCAATTTGA	3575
1272	AlaGlnPheValAlaLeuAsnValIcLysAlaValSerGlnHisLysLe	1288
3576	CCGCGGTG...TTTGGCGAAGACCGCGCCGACAGCGMTTGGACAAGCA	3622
1288	rgLysLeuGlnMetAsnAsnGluGlyGlnThrAsnValThrPalSerAsnT	1305
3623	TCCGGACACCAACACATACCGTTGGCAAGATTTCGCGGCTACCGCCAA	3672
1305	hSerMetAsnLysAsnLysSerSerSerGlnThrAlaArgPheSerSer	1321
3673	CAACGACCTTGGCGCAATCGGTATGGAGAAAACTCGGACAGCGG...	3720
1322	LysSerThrGlnThrGlnLeuGlyThrAspSerIleThrIleSerAsnAsnVal	1338
3721	CGCGTGGGATCCGTGTTTGGCACACACCGGACGGAACACTTCGACG	3769
1338	IcLysLeuGlyGlyValPheThrThrValArgAsnSerAsnAsnProPheAsp	1355

```

3770 ACGGCGATCGCAACTCGGACGGCTTGGCCACGGCGCGTTTGGGGCAA 3819
      ::::::::::::::::::::
1355 ysAlaThrSerLysAsnThrLeuAlaValAsnPhenylSerLysTyr 1371
      ::::::::::::::::::::
3820 TACGGCATCGGACGAGTTCGACATCGGCATCAGCACGGCGGGGCTTTAG 3869
      ::::::::::::::::::::
1372 TyrAlaAspAsnHisThrPtyrLeuGlyIleAspLeuGlyTyrGlyLysPhe 1388
      ::::::::::::::::::::
3870 CAGCGGCANTCTTTCAGACGCGATCGGAGCAAAATCCGCGCGCGTGC 3919
      ::::::::::::::::::::
1388 eGlnSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThr 1405
      ::::::::::::::::::::
3920 TGCATTACGGCATTCAGCAGCATACCGCGCGGCTTTCGGCGGATCGGC 3969
      ::::::::::::::::::::
1405 IagLInPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyLysPheGly 1421
      ::::::::::::::::::::
3970 ATCGAAACGCTACATCGGCGCAACGGCGATTTGTCGCAAAAGCGGATTA 4019
      ::::::::::::::::::::
1422 ILeuThrProIleValGlyAlaArgTyrSerTyrLeuSerAsnAlaAsp 1438
      ::::::::::::::::::::
4020 CCGCTACGAAACGCTCAATATCGCCACCCCGCTCTGCTTCACACCGNT 4069
      ::::::::::::::::::::
1438 eAlaLeuAspGlnAlaAlaArgIleLysValAsnProIleSerValLysThr 1455
      ::::::::::::::::::::
4070 ACCGACGCGGCATTAAGCAGATTAATTCATTCACACCGCGCACACATN 4119
      ::::::::::::::::::::
1455 IAPheAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyLys...Phe 1470
      ::::::::::::::::::::
4120 TGCATACACGCTTATTNNAGCCTGCTCCTACCGGATGCGCGTGGCGAA 4169
      ::::::::::::::::::::
1471 SerValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySe 1486
      ::::::::::::::::::::
4170 ACTCGCAACCGCGCTCAATATCCGCGTATGCGTCAGGATTTTCGCAAAA 4219
      ::::::::::::::::::::
1486 rGlyLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsn 1503
      ::::::::::::::::::::
4220 CCGCGATCGCGAATGCGCGCTAAACCGCGAATCAAGGTTTCACGCTG 4269
      ::::::::::::::::::::
1503 IngLInGlnIleTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeu 1519
      ::::::::::::::::::::
4270 TCCNTCCACGATCGCGCGCGCAAGCGCAACTGGAGCGCACACACAG 4319
      ::::::::::::::::::::
1520 SerLeuIleGlyLeuThrLysAlaLysGlnAlaGlyLysGlnLysThr 1536
      ::::::::::::::::::::
4320 CCGCGGCATCAAAATTAGGCTAC 4341
      ::::::::::::::::::::
1536 rAlaGlnLeuLysLeuSerPhe 1543
      ::::::::::::::::::::
seq_name: pir2:S61314

```

```

seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HF13
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61314
R:Limholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria
A:Reference number: S61314; MUID:95302961
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: EMBL:X82474; NID:g732873; PIDN:CAAS7857.1; PID:g732874
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
Quality: 1551.50      Length: 1754
Ratio: 1.730          Gaps: 51

```

Percent Similarity: 51.140 Percent Identity: 26.910

alignment_block:

US-09-303-518D-651 x S61314 ..

Align seg 1/1 to: S61314 from: 1 to: 1561

```

49 AAACCGCGCGCATCCGCTTCGCGCTGCTACTTACGATATGCCGTC 98
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
      ::::::::::::::::::::
99 GTTCGGCATCTTCCCAAGCTTGGCGGACACACTTATTTTCGGCATCA 148
      ::::::::::::::::::::
18 eTyrAlaLeuThrProTyrSerGlnAlaAlaLeuValaAspAspVala 35
      ::::::::::::::::::::
149 ACTACCAATACATACCGGACTTTCGCGAAATTAAGCAATTTCGAGTC 198
      ::::::::::::::::::::
35 sPtyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
      ::::::::::::::::::::
199 GGGCGCAAGATATTTGAGTTCACAAACAAAGGGGAGTTCGCGCAA 248
      ::::::::::::::::::::
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleLys 68
      ::::::::::::::::::::
249 ATCAATGCAAAAGCCCGCATGATGATTTTCTGTGCTGCGCTAAC. 297
      ::::::::::::::::::::
68 nAlaLeuSerAsnValProMetIleAspPheSerValaAlaAspValAsnL 85
      ::::::::::::::::::::
298 .GGCGTGGCGCATTTGCTGGCGATCATATTTGTGACCGTGGCAT 345
      ::::::::::::::::::::
85 ysArgIleAlaThrValValaAspProGlnTyrAlaValSerValLysHis 101
      ::::::::::::::::::::
346 .....AACGGCGCTATPACACGTTGA 368
      ::::::::::::::::::::
102 AlAlaLysAlaGlnValHisThrPheTyrTyrGlyGlnTyrAsnGlyHis 118
      ::::::::::::::::::::
369 TTTTGTGCGGAGGAAGNAATCCGATCAGACCGCTTTTCTTACCAA 418
      ::::::::::::::::::::
118 nAspValAlaAspLysGluAsn.....GluTyrArgY 129
      ::::::::::::::::::::
419 TTGTGAAAGAAATATTAATTAAGCCTGACAAATTCACACCTTACAAAGCG 468
      ::::::::::::::::::::
129 aValaGlnGlnAsnAsnTyrGlnProHisLysAlaTrpSerAlaSerAsn 145
      ::::::::::::::::::::
469 .....GATTNCAATATCCGCGTTCGATTAATTTGTCAC 503
      ::::::::::::::::::::
146 LeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLysPheValTh 162
      ::::::::::::::::::::
504 AGATGCAAGACCTGCGAAATGACGAGTGCACATGAGGGG...AATACCT 550
      ::::::::::::::::::::
162 rGluValAlaProIleAlaProThrAspAlaGlyGlyLeuAspPThr 179
      ::::::::::::::::::::
551 ATTCCGATTAAGAAATATCCGAGCGTGTCCGCAATCGGCTCAGGACAC 600
      ::::::::::::::::::::
179 yLysAspLysAsnArgPheSerSerPheValaValaGlyAlaGlyArg 195
      ::::::::::::::::::::
601 CACTAT.....TGGCGTATGATGATGCAAAACGCG 632
      ::::::::::::::::::::
196 GlnLeuValTyrGluLysGlyAlaTyrHisGlnGluLysGlnLysGln 212
      ::::::::::::::::::::
633 C.....GATTATCCCTACTCCGCGGCAATGTTAATGGCGGC. 669
      ::::::::::::::::::::
212 yTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleAlaGlyT 229
      ::::::::::::::::::::
670 .....AATACACAT..ATG 681
      ::::::::::::::::::::
229 hrProTyrLysAspIleAsnIleAspGlnThrMetAsnThrGlnGlyLeu 245
      ::::::::::::::::::::
682 CAGGTTGGGCAATATGCGGTANTAGTTTGAGCGCGCGGATGAGGCCA 731
      ::::::::::::::::::::
246 ILeuGlyPheGlyAsnHisAsnThrHisTyrSerAlaGlnGlyLeuLysG 262
      ::::::::::::::::::::
732 TGCCACGACTATGCGCCTATGCGGATTCAGAGTGGCGGACGAGCGAGCG 781
      ::::::::::::::::::::

```



```

1295 hrglyAlaSerLeuThrArgHisIleAlaGlnLysSerArgAlaAspAla 1311
3601 ..... AACGGNTTGGACAAGCGCATCGCGAACCACCAACCTACCG 3644
1312 GlnLysSerValTyrMetSerAsnIleLysTyrGlnArgAspTyrAl 1338
3645 TTCCAGATTTCCGCGCTACCGCCCAACACCGACCTGCGCAATTCG 3694
1328 aSerAlaGlnTyrArgArgPheSerSerLysArgThrGlnThrGlnIleG 1345
3655 GTATGACAAAACCTCGCGACGCGG...CGGTCGCGATCTGTTTCG 3741
1345 LysLeuAspArgSerLeuSerGlnLysMetGlnIleGlyValLeuThr 1361
3742 CACAAACCGGACGAAACANCTTCAGACGGGATTCGGCACTCGGACG 3791
1362 TyrSerAspSerGlnHisThrPheAspGlnAlaSerGlyLysAsnThr 1377
3792 GCTTGCCACGCGCGCGCTTTTCGGCAATACGGCATCGGC...AGTTTCG 3838
1378 .PheValGlnAlaAsnLeuTyrGlyLysTyrLysLeuAsnAspAlaTyr 1394
3839 ACATCGGATACGACGCGCGCGGCTTTTACGACGGCATCTTTCAGAC 3888
1394 yValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuGlnThr 1410
3889 GGCAATCGGACGAAATCCGCGCGCGGTGCGATACGGCATTCAGGCG 3938
1411 GlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuThr 1427
3939 ACGATACCGCGCGGTTTCGCGGATTCGCGCATGCAACCGTCATCGCG 3988
1427 uGlyAsnThrLeuLysIleAsnGlnPheGlnIleValProSerAlaGly 1444
3989 CAACGCGTATTCGTCACAAAACGGGATTCGCGCATGCAAAAGCTAAT 4038
1444 leArgTyrSerArgLysSerAlaAspTyrLysLeuGlyAsnAspSer 1460
4039 ATCCGACCCCGGCTGCTTCACCGGNTACCGGCGGCGCATTAAGGC 4088
1461 ValLysValSerSerMetSerValLysThrLeuThrAlaGlyLeuAsp 1477
4089 AGATTATTCATCAACCGCGCAACATTCATCAACNCTTATTATTA 4138
1477 eAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProLeuAs 1493
4139 GCCGTCTCTAT...ACGATGCGCGCTTCGCGCAAGTCCGACACGCGTC 4185
1493 eAlaAlaTyrPheAlaAsnTyrGlyLysGlyValAlaAsnValGly 1509
4186 AATACCGGNTATGGCTCAGATTTTCGCAAAACCGGACAGTGGGATG 4235
1510 AsnSerPheAlaTyrLysAlaAsnGlnGlnIleTyrSerAla.... 1524
4236 GGGCGTAAACCGCAATCAAGGTTTCGCGCTGCTCCATCCGCGCGG 4285
1525 .GlyAlaAlaLeuLeuTyrArgAsnValThrLeuAsnValAlaSer 1541
4286 CCGCAAGGCGCACTGGAACGCGACACACGCGCGGCGATCAATTA 4335
1541 leThrLysGlyLysGlnLeuGlnLysSerGlyGlnIleLysIle 1557
4336 GGCTACGCGCTGG 4347
1558 GlnIleArgPhe 1561

```

seq_name: p12:A37023

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
 N/Alternate names: immunoglobulin A1 protease type 1
 C/Species: Haemophilus influenzae
 C/Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000

C/Accession: A37023
 R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Kilian, M.
 Infect. Immun. 57, 3097-3105, 1989
 A/Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophilus influenzae
 A/Reference number: A37023; MUID:89379374
 A/Accession: A37023
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1541 <POU>
 A/Cross-references: GB:X64357; MID:g43560; PIDN:CAA45708.1; PID:g43561
 A/Experimental source: serotype b
 C/Superfamily: IgA-specific metalloendopeptidase
 C/Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1536.50 Length: 1722
 Ratio: 1.677 Gaps: 56
 Percent similarity: 53.194 Percent identity: 26.887

alignment_block:
 US-09-303-518D-651 x A37023 ..

Align seg 1/1 to: A37023 from: 1 to: 1541

```

64 CGCTCTCGCGCTGCTTACGATATGCTGTCGTCGCGATTCCTCC 113
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAAGCTTGGCGGACACACTTATTCGCGATCACTCACTACATATC 163
21 oTyrThrGlnAlaAlaLeuValArgAspValAspTyrGlnIlePhe 38
164 GCGACTTGGCGGAAATTAAGGCAAGTTTCAGTTCGCGCGGCGAAAGTAT 213
38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTNTACCAACAAAGGAGGAGTGTGTCGCGCAATGACAAACG 263
55 LeuValLysAspLysAsnLysAsnLysAspLeuGlyThrAlaLeuProAsn 71
264 C...CCGATGATTTGATTTTCTGTCGTCG...CGTAAACGCGTGGCG 307
71 yIleProMetIleAspPheSerValAlaValAspValAspLysArgIleAla 88
308 CATTCGCGGCGCATCAATATTCGACGTCGCGCAAT...AACGCGCG 354
88 hLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATTAACAGCTGATTTTGGT.....GCGGAAGGAGAGNA 389
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
390 TCCCGATCAGCAGCT.....TTTCTTACCAATTCG 421
121 nAlaLysAlaIleArgAspValSerSerGlnIleAsnArgTyrPheSer 138
422 TGAAGAATAATATAT.....AAG 441
138 alGlnLysAsnGlnTyrProThrLysLeuAsnGlyLysThrValThrThr 154
442 CCTGACAAATCAGACCCCTTACAGCGGATTAACATATTCGCGGTTTGA 491
155 GlnAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgLeuAs 171
492 TAAATTTGTCAGATGACAGAACCTGTCGA...ATGACAGTGCATGTA 538
171 pLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaSerSer 188
539 GGGGGAATACCTTTCGATTAAGAAATATTCGCGAGGTGTCGCGATC 588
188 sPAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204

```

```

589 GGCTCAGACACACTATTGGCGTTATGATGATGACAAACGCGGATTT 638
      ||||| :|||
205 GlySerGlySerGlnPheIleTyr.....LysGlyGlyAsp... 216
      :|||
639 ATCCCTCCCGGCGCATGGTTAATGGCGGCATACATATACGAGGTT 688
      :|||
217 .AsnYrSer.....LeuIleuAsnAsnHisGluValGly. 228
      :|||
689 GGGGAATATAT.....CGCGTANTT 708
      |||||
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
      :|||
709 AGTTTGAGCGCGCATGTCCGCATGCCAAGCAG.....TATGG 746
      :|||
245 GlyThrProTyrLysValAsnHisGluAsnGlyLeuIleGlyPheG1 261
      :|||
747 C.....C 748
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278
749 CTATGCCGATTCGAGTGGCGGAGGAGACGCGTTCCGCATGTTTAT 798
      :|||
278 roleuThrAsnTyrAlaValIleuGlyAspSerGlySerProLeuPheVal 294
799 TATGACAAACAAACATATAATGGCTGCTCAACGAGATTTTACAAACCG 848
      :|||
295 TyrAspArgGluLysGlyLysThrLeuPheLeuGly.....Se 307
849 CTACCCCTTATCCGCGAGGAAACGGTTTCCAGCTGATACGCAAGAT 898
      :|||
307 rTyrAspPheThrPalaGlyTyrAsn.....LysLysSerT 319
899 GGTCTACGATGACATTTACAGAGCGATACACATACCGCTNTTGTAA 948
      :|||
319 rPoleGlnLutrrpAsnIleTyrLysSer..... 327
949 CCGCGCAGTACGAGACATTTTCTTTACATCC.....AACACAA 989
      :|||
328 .....GlnPheThrLysAspValIleuAsnLysAs 337
990 CGGTACGGGTAGCGTACAGAAACCAACGAAAGTNTCC...AATCCA 1036
      :|||
337 pSerAlaGlySerLeuIleGlySerLysThrAspTyrSerThrPheSera 354
1037 ACCTTAAGTACAGACAGTCCGACTGTTTGACGAATCTTTGATTAAC 1086
      :|||
354 snGlyLysThrSerThrIleThrGlyGlyLysSerLeuAsn..... 368
1087 GATTAAGAACAGTTTACGCGGAGGGGTGTTAATCAGTACCGTCCAG 1136
      :|||
369 .....ValAspLeuAlaAspGly.....LysAspLys 377
1137 GTTAACAACGCGTGAACCTTTCTTTATGATTAACGCAACGCGCAAC 1186
      :|||
377 pProAsnHisGlyLysSerValThrPhe.....GluGlySerGlyThrL 392
1187 TCATCTTATCAACAACATCAACAGAGCGCGCGGCTTTGATTTGAA 1236
      :|||
392 euThrLeuAsnAsnHisLeuAspGlnGlyAlaGlyGlyLeuPheGlu 408
1237 GGTGATTTTACGGTC...TCGCGTAAACAAACGAAACGTGGCAAGCGC 1283
      :|||
409 GlyAspTyrGluValLysGlyThrSerAspAsnThrThrPlyGlyAla 425
1284 GGGCGTTCATATCAGTGAAGACGATCTTACTTGGAAAGTAAAGCGCG 1333
      :|||
425 aGlyValSerValAlaGluGlyLysThrValThrTrpLysValHisAsn 442
1334 TGGCAACAGCGCGCTGTCAAAATCGCAAGGAGCGCTGACGATTTCAA 1383
      :|||
442 roGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 458
1384 GCCAAAGGGGAAACCAAGGCTCGATCAGCGTGGCGAGCGGTACATCAT 1433
      :|||
      :|||
459 GlyThrGlyAspAsnLysGlySerLeuLysValGlyLysPoleGlyThrValI1 475
1434 TTGGATGACGAGGCAACGATTAAGGCAAAAAACGCTTAGTGA 1483
      :|||
475 eLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheLeuLaserV 491
1484 TCGGCTTGNTACCGCGCAGCGGTACGGTGCACATGTAATGCCATATCAG 1533
      :|||
491 aGlyIleValSerGlyArgSerThrLeuValIleuAsnAspAspLysGln 507
1534 TTCAACCCCGCAACAATCTATTTCCGCTTCCGCGCGGAGCGTTGGATTT 1583
      :|||
508 ValAspProAsnSerIleTyrPheGlyPheArgGlyGlyLysLeuAspLe 524
1584 AAAGGGCATTCGCTTCCGTTCCACCGTATTTCAAAATACGATGAAGGG 1633
      :|||
524 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspAspGly 541
1634 CGATGATTCGNCNATCATATATGCCACACACATCCACGTTACCATTACA 1683
      :|||
541 laArgLeuValAsnHisAsnMetThrAsnAlaSerAsnIleThrIleThr 557
1684 GCGATGAAGATATTACACAACCGAGT..... 1710
558 GlyLysSerLeuIleThrAspProAsnThrIleThrProTyrAsnIleAs 574
1710 ..... 1710
574 palAProAspGluAspAsnProTyrAlaPheArgArgIleLysAspGlyG 591
1710 ..... 1710
591 LysGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyrTyrAlaLeuArgLys 607
1710 ..... 1710
608 GlyAlaSerThrArgSerGluLeuProLysAsnSerGlyGluSerAsnG1 624
1711 .....GSTAGAAAT.....A 1720
624 uAsnThrPLeuTyrMetGlyLysThrSerAspGluAlaLysArgAsnValK 641
1721 TCATATGACTTAATTACAGCAAGAAATTCCTACACAGGTTGTTGGC 1770
      :|||
641 etAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly 657
1771 GAGAAAGATACGACCAACGAAACGGGCGGCTCAACCTGTTTACAGCC 1820
      :|||
658 GluGlnGlu...GlyLysAsnAsnGlyAsnLeuAsnValThrPheLysG1 673
1821 CGCGCGAGGACCGCGCCCGCTTCCGCGGAGCAACAAATTTTAAG 1870
      :|||
673 LysSerGlnGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 690
1871 GCAACATCAGCAACAAACGCAAACTGTTTTCAGCGGAGCGAGCGAGCA 1920
      :|||
690 LysAspLeuThrValGluLysGlyThrLeuPheLeuSerGlyArgProThr 706
1921 CCGCAGCGCTTACATCATTTAGAGAGCGGCTGCTCAAAATGGAAGGT.. 1968
      :|||
707 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspProH1 723
1969 ....ATCCCAAGGAGGAATGCTGGGAGCAACGATCGGATNACCGCA 2014
      :|||
723 sPheAlaGluAsnAsnGluValValGluAspAspThrPheAsnArg 740
2015 CGTTTAAAGCGGAAATTTCCATATTTCAGGCGGAGCGGCGGTGATTTCC 2064
      :|||
740 snPheLysAlaThrThrMetAsnValThrGlyLysAsnAlaSerLeuTyrSer 756
2065 ...CGCATGTTGCCAAAGTGGAAGCGGATTCGATTTGACATACATCAGC 2111
      :|||

```

```
757 GlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAsnLysAl 773
2112 CCAAGCAGTTTGTGTGTGGACCCGATCAAGCCATACATCTGTACAC 2161
773 aglnvalhlsllleglytyr.....lysthrglyasprhvalcysvala 788
2162 GTTCGGACGTGGACNGSTGTGACAAATGTGTGGANNAANCAATTCGAC 2211
788 rgsersprythrghlytyrvalthrCysThrThraspLysleuser... 803
2212 GATTAAGTATGCTCATTCATTAAGACNGACNTNAGCGGCANTGTNAG 2261
804 AspLysAlaLeuAsnSerPheAsnProThr..... 813
2262 NCTNNCCNATNAGNTNNTTNAANCTCNCNGGCGNTGNCNCCTNAAAG 2311
813 ..... 813
2312 GCATCTTGTGCAATAGCGGATACAGCTTATACAGTCAGCCACAAGCC 2361
814 ..AsnLeu..... 815
2362 ACCCAAAAGCGACCTTACCTCGTGGGCAATGCCCAAGCAACATTTAA 2411
816 .....ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheValLeuGl 830
2412 TCAGCCACATTTAAAGCGCAACNCATCGGNTTGGGCAATGCTTCATTTA 2461
830 ylysAlaAsnLeuPheGlyThrIleGlnSerArgLysAsnSerGlnVala 847
2462 ACTTAAGCAACAAGCCCGCACAAAAGCGACGTCTGACGCTTTCGACAA 2511
847 rgleuthrGluAsn..... 851
2512 GCTAAGCAAAAGTAAGCATTCGCGACTCAACGCGCAATGTCTCCTAGC 2561
851 ..... 851
2562 CGATAGCGAGTATTCATTTTGAACAAGCGCTTACCGGACAATCA 2611
851 ..... 851
2612 GCGGACGACGAGNACAGCATTTACACTTAAAGACAGCAATGCACGCTG 2661
852 .....SerHisIleThrPheLeu 856
2662 CCGTGAGCAGCGAATTAGCAATTTAAACCTTGACAGCCACCATAC 2711
857 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleHis 873
2712 ACTCAATTCGCGCTATCGCCACAGATGTCGACGCGCGCAAAACGCGCAG 2761
873 sleuAsnSerAlaAspAsnSerAsn..... 881
2762 TGTGACACAGCGCGCGCGCTTCCCTTATTCGCTTACA 2811
882 .....AsnValThr 884
2812 CCGCCACCTGGTAGAATCCGTTTCAACAGCTGACGTAACGGGAA 2861
885 .....LysTyrAsnThrIleuThrValAsn...Se 893
2862 ATTGACAGCTCAAGACATTCGCTTATGTGGAACCTTCGCGTACC 2911
893 rleuserGlyAsnGlySerPheTyrLeuThrAspLeuSerAsnLysG 910
2912 GAAGGACAAATTGAAGCTGGCGGAATTCGGAAGNACTTAACTTACCTG 2961
910 lnglyAspLysValValThrLysSerAlaThrGlyAsnPheThrLeu 926
2962 GCGGTCAACATACCGGACGAAACGCTGATCAATTCAGAGGT 3011
927 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrle 942
3012 AGTGAAGGGAAGAACAAACACCGCTGTCCGAAACCTTAATTTCACC 3061
942 upheaspAla.....SerLysAlaGlnArgAspHisIleAsnValSerL 957
3062 TGCAAAAGACACGTCGATCGCGCGGCGGCTTACCAATCATTCGCG 3111
957 euValGlyAsnThrValAspLeuGlyAlaThrPlystyryLysLeuArgAsn 973
3112 AAAGACGCGAGTTCCGCTGACATATCCGTCAAAGACAA..... 3153
974 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGl 990
3153 ..... 3153
990 nThrValAspThrThrAsnIleThrProAsnAsnIleGlnAlaAspV 1007
3154 .....GAGCTTTCGACAAACTCGCGCAG..... 3177
1007 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1023
3177 ..... 3177
1024 ProValProProAlaProAlaThrProSerGluThrThrGluThrVa 1040
3178 .GCAGAACCCAAAACAGCGCGGAAAA.....GACACGCGCAAA 3217
1040 lAlaGlnAsnSerLysGlnGluSerLysThrValGlnLysAsnGluGlna 1057
3218 GCGTTGACGGCGTATTCGCGCGCGCGATGCGCGGAA..... 3258
1057 spAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073
3259 .....AAGACGAAAGCGTTGCCACACCGGCGCG 3287
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087
3288 GCNCGCAGCGCGGGAATTCGGCATTTATCAGCGCGAG..... 3327
1088 ...SerGlySerLysThrLysGluThrGlnThrThrGluThrLysGluT 1103
3328 .....GAAGAGAAAAAGCGGTGCAGCGGATTAAGAC 3360
1103 hralaThrValGluLysGluGluLysAlaLysValGluThrGluLysThr 1119
3361 AGCGCNTTGGCGAAA.....CAGCGGAGCGGAAACCGCGCGG 3401
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluInSe 1136
3402 TACACCGCTTCCCGCGCGCGCGCGCGCGGATTTGCG..... 3447
1136 rGluThrValGlnProGlnAlaGluProAlaArgGluAsnAspProThrV 1153
3448 .....CAACGCGACGCCAA..... 3462
1153 alaSnIleLysGluProGlnSerGlnThrAsnThrAlaAspThrGlu 1169
3463 .....CCGCAACT..... 3471
1170 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluSe 1186
3471 ..... 3471
1186 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1203
3472 .....CAACC.....CAACG 3483
1203 hrProAlaThrThrGlnProThrValAsnSerGluSerAsnLysPro 1219
3484 CAACGC..... 3489
1220 LysAsnArgHisArgArgSerValArgSerValProHisAsnValGluPr 1236
```

```

3490 .....GACCTGA 3496
1236 oalathrthrsersasnspargserrthvalalaleucysaspleur 1533
3497 TNAGCGGTTAGCAATGAGTGGATTTCCGCC ..... 3537
1253 hrserthrasnthrasnlaivalaleuaseraspalaalvalalvalagln 1269
3538 .....ACGCTCAACAGCGTTTCCGCTACAGACGAGCAAAATGGACCGCT 3581
1270 Phevalalaleuasnvalgllysalavalserglnhstileserglnle 1286
3582 G...TTTCCGCAAGACCGCCCAACGCGTGGACAAACGACATCCGGN 3628
1286 uglmethasnsngllyglnlythrasnvaltrpvalserasnthrserm 1303
3629 ACACCAACATACCGCTTCCGCAAGATTTCCGCTACCGCCCAACAAAC 3678
1303 etasnlyasnthyserserserglntrargargpheserSerlyseSer 1319
3679 GACCTGCGCCAAATCGGTATGCAGAAAACCTCGGACGCGG...CGCGT 3725
1320 Thrghnthrglnleuglytrpaspdlnthrilesersasnsvalglnle 1336
3726 CGGCATCTGTTTCCGCAACCGGACGCAAAACANTTCCGACGACGCA 3775
1336 ucllyglvalpetheritytrvalargasnserasnsnphesaplysalat 1353
3776 TCGGCAATCGGACGCGCTTCCGCAACCGGCGCTTTTCCGCAATACGC 3825
1353 hrserlyasnthrasnlaivalasnphetyrserlytyrtyrtyrtyr 1369
3826 ATCGGACGATTCGACATCGGATCGACGCGGCGGCTTTAGACGCG 3875
1370 Aspasnhtstrptrytleuglylleaspleucllytyrlyllyspheglnse 1386
3876 CANTCTNTCAGACGCGATCGGACGCAAAATCCGCGCGCTGCTGCATTT 3925
1386 trysleuclnthrlnhstasnsalalysrphelalarghtsthralaglnp 1403
3926 ACGGATTCAGGACGATACCGCGCGGCTTCCGCGGATTCGGGATTCGAA 3975
1403 heglyleuthralaglylalsalaphasnleuglyllyasnphedlyllethr 1419
3976 CCGTACATCGGCGCAACGCGATTTCCGCAAAAGCGATTCACCGCTA 4025
1420 proillevalgllyvalargtyrserlytleuaserasnsalalasphealale 1436
4026 CGAAACGCTCAATATCGCCACCGCGCTTCCGCTTCAACCGGATACGCG 4075
1436 uasprglnalargllyllyllysalasnproilaservallysthralapha 1453
4076 CGGGGATTAAGCAGATTAATTCATCAACCGGCGCAACACATNTCCATC 4125
1453 lagnlvalaspleuaserlytrnlythrasnleuglylu...PheSerVal 1468
4126 ACNCGTTATTTNAGCGCTGCTTATACCGATGCCGCTTCCGCGCAAGTCCG 4175
1469 thrproilleuaseralargtyr...AspalaasnnglnlySerGlyly 1484
4176 AACAGCGCTCAATACCGCGCTTGGCTCAGATTTCCGCAAAACCGCA 4225
1484 stllesnvalasnnglytyrasprrhealalatytrasnvalglubasnnglnng 1501
4226 GTGGCGAATGGGCGTAACGCGCAATCAAGGTTTCCGCTGCTCCATC 4275
1501 InlythrasnlaaglyleuylsleuylslythrasnvallylsleuSerleu 1517
4276 CACGCTGCGCGCGCAAGGCGCAATCAAGGCGCAACACGCGCGG 4325
1518 lileglyllyleuthrasnlaalysglnalaglnlyllysthrAlagI 1534
4326 CATCAATTAAGGCTAC 4341

```

```

1534 uleuylsleuSerPhe 1539
seq_name: plr2:A81018

```

```

seq_documentation_block:

```

```

serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogr
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81018
R:Reteljin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rt, H.; Qin, H.; Yamathewan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizze, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: A81018
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-1431 <TEXT>
A:Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAF42325.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1998

```

```

alignment_scores:

```

```

Quality: 1522.00 Length: 1639
Ratio: 1.770 Gaps: 51
Percent Similarity: 52.471 Percent Identity: 27.639

```

```

alignment_block:

```

```

US-09-303-518D-651 x A81018 ..
Align seg 1/1 to: A81018 from: 1 to: 1431

```

```

55 GGGCGGATCGGCTTCCGCGCTTATAGCCATATGCTGCTTCCG 104
17 Glyleuphealavalserproalatyrserserilevalargasn..... 31
105 CATTCTTCCCAAGCTTGGGCGGACACACTTTTCGGCATCAATAC 154
32 .....AspValaspTyrg 36
155 AATATATCGGACTTTCGCGAAATTAAGCAATTTGCGATCGGCGG 204
36 Inlyrpheargasprrhealaglualsnlysllyalapherthralal 52
205 AAGATATTTAGGTTNTRCAACAAAAAGGAGTGTGTCGCAATCAT 254
53 Serasnllleserlleaglinalasplysglnlyllylleucllyalvalle 69
255 GACAAAGCCCGGATGATGATTTTCTGCGTGCCT...AACGGCG 301
69 uasnnglyllepromerproasprheargvalaserasnarglnthralal 86
302 TGGCGGATTTGGGCGGATCAATATATTTGAGCGTGCACATACGG 351
86 lialatrleuvalinhstproglntytrvalasnservallyllysthrasnsval 102
352 GCGTATTAACAACGTTGATTTTGGCGGGAAGAAATATATTAAGCTGACAT 401
103 glytyrlyserlleglinphelgylasnaspThrclnlnsnprogluugl 119
402 CCGTTTCTTACCAATTTGCAAAAGAAATATATTAAGCTGACAT 451
119 palatytrthrtarglvalserarg.....Asp 130
452 CACACCGCTTCAACGCGATTAATATATGCGGCTTGCATTAATTTGTC 501
130 rohtspro...AsptyrasptyrthstleuProatrgleuasnlyleuval 145

```



```

3573 GAACCGGCTGTTTCCGAAACCGCGCAACGCGTGTGGACAGCNGCA 3622
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1173 eAspArgHisLeuThrAspProGlnGlnAsnIleTyrPheGlnThrG 1190
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3623 TCCGGNACCAACACATACCGTTCGCAAGATTTCGGCGCTACGCCCA 3672
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1190 LThrGlnGlnThrAspTyrHisSerGlyThrHisArgProGlyGln 1206
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3673 CAACACCGCTGGCCAAATCGGTATGCAAAAACCTCGGACGGCGG... 3720
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1207 ThrThrAsnTyrAlaHisIleGlyIleGlnThrGlyLeuThrAsp 1223
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3721 .CGCGTCGGATCCCTGTTTCCGACACCGGACCGCAAAACACTTCGAG 3769
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1223 uSerValGlyThrIleLeuThrAspGlnArgThrAsnAsnArgPheAs 1240
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3770 ACGGCATCGGC.....AACTGGCGACGGCTGGCCACGGCGCGTTTC 3813
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
1240 LuGlyValSerAlaArgAsnArgSerAsnGlyAlaHisLeuPheVal 1256
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3814 GGGCAATACGGC.....ATCGGACAGTTCGACATCGGATCGGACAAATCC 3857
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1257 GlyGlnAsnGlyAlaLeuPheAlaAlaAspLeuGlyTyrSerAsn... 1272
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3858 CGCGGATTTCAGCGCGGACANTCTNTCAGACGCGATCGGAGCGCAATCC 3907
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1273 .....SerArgThrArgPheThrAspTyrAspGlyAlaAlaValA 1286
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3908 GCGCGCGCGTGGCTGCATTCAGCGCATTCAGGACGATACGGCGCGTTTC 3957
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1286 rGArgHisAlaTyrAspAlaGlyIleAsnThrGlyLeuHisLeasp... 1301
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3958 GCGGATTCGGCATCGAACCGTACATCGC..... 3987
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1302 ThrGlyIleAsnLeuArgProTyrAlaGlyIleArgIleAsnArgSerAs 1318
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
3988 .GCAACGCGCTATTCGTCGCAAAACGGGATTCAGCGTACGAAACGTCA 4036
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
1318 nGlyAsnArgTyrValIleuAspGlyAlaGln..... 1328
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4037 ATATCGCCACCGCGGCTGGTTCACCGTACCGTACCGCGCGCATTAAG 4086
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1329 ..IleAsnSerProAlaGlnIleGlnThrThrThrHisAlaGlyIleArg 1344
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4087 GGAATTTATTCATTCACACCGCGGACACATTCACATTCACATTCATTT 4136
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1345 LeuAspLysThrValGlnLeuGlyGln...AlaLysLeuThrProAla 1360
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4137 NAGCCGTGCTAT.....ACCGATGCGCGTTCGGGCAAGTCCGAAAC 4180
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1360 eSerSerAspTyrTyrHisThrArgIleAsnSerGlySerAlaLeuSer. 1376
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4181 GCGTCAATTCGCGNGTATTCGCTCAGGATTCGCGCAAAACCGCGAGTGG 4230
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1377 ..ValAsnAspArgTyrThrLeuLeuGlnGlnAlaHisGlyThrLeuHis 1392
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4231 GAATGGGCGGTAACCGCGAAATCAAAAGTTTCACGCTGCTCCATTCAC 4280
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
1393 ThrLeuGlnIleAspAlaGlyTyrLysGlyTyrAsnAlaLysLeuHisAl 1409
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4281 TGGCGCGCGCAAGNCCGCAACGTGGAAAGCGCAACAGCGCGGATCA 4330
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1409 aAlaTyrGlyLysAspSerAsnThrAlaArgHisLysGlnAlaGlyIle 1426
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
4331 AATTAGCTACGCTGG 4347
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
1426 ySileGlyTyrTrpAsnTrp 1431

```

seq_name: p1r2:H64106

seq_documentation_block:

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain N/Alternate names: immunoglobulin A1 protease type 1

C:Species: Haemophilus influenzae
 A:Variety: strain Rd KW20
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
 C:Accession: H64106; A41500
 R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Meldna
 , D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venti
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; M01D:95350630
 A:Accession: H64106
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1694 <YIGR>
 A:Cross-references: GB:032779; GB:I42023; MID:91574009; PIDN:AMC22651.1; PID:9157401
 A:Experimental source: strain Rd KW20
 R:Grundy, F.J.; Plaut, A.G.; Wright, A.
 Infect. Immun. 58, 320-331, 1990
 A:Title: Localization of the cleavage site specificity determinant of Haemophilus in
 A:Reference number: A41500; M01D:90129281
 A:Accession: A41500
 A:Molecule type: DNA
 A:Residues: 1-377 <GRU>
 A:Cross-references: GB:X59800
 A:Experimental source: strain Rd KW20
 A:Note: the authors translated the codon TGG for residue 319 as Thr
 C:Function:
 A:Description: this protease is classified as type 1 because it cleaves at a prolit
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloprotease

alignment_scores:
 Quality: 1519.00 Length: 1873
 Ratio: 1.653 Gaps: 54
 Percent Similarity: 49.066 Percent Identity: 24.826

alignment_block:

US-09-303-518D-651 x H64106 ..

Align seg 1/1 to: H64106 from: 1 to: 1694

```

64 CGCTTCGCGCTGCTAGCATATGCGTGGCTGGCATTCCTCC 113
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 LysPheLysLeuAsnProIleAlaLeuThrValAlaTyrAlaLeuThr 21
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 CAAAGCTTGGCGGACACACTTATTCGGCATCACTACCAATATCATC 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
21 oTyrThrGlnAlaIleuValAlaArgAspArgValAspTyrGlnIlePhe 38
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
164 GCGACTTGGCGGAAATTAAGGCAAGTTTCGACGTGGGCGGAAAGATAT 213
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
38 rGAspPheAlaGlnAsnLysGlyArgPheSerValGlyAlaTyrAsnVal 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 GAGTNTACAAACAAAAGGAGGAGTTCGCGCAAAATCATCATGAAC 263
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 GluValAlaArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsn 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 C...CCGATGATGATTTTCTGTGTGCG...CGTAACGGCGTGGCGG 307
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 yIleProMetIleAspPheSerValValAspValAspLysPheArgIleAla 88
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
308 CATTCGTCGGCGGATCATATATTCGTGCGCTGGCAGCAT...AACGGCG 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 TATAACAACGCTGATTTTGGT.....GCGGAGGAGAGGMA 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
390 TCCCGATCAGCACCGCT.....TTTTCTTACCAAAATTG 421
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
121 nAlaLysSerHisArgSpValSerSerGluGluAsnArgTyrPheSerV 138
422 TGAAGAAGAAATATAT.....AAG 441
138 aIGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
442 CCTGACAAATTCACACCCCTTACACGGCGATTATGCATATGCCGCTTGA 491
155 GluAspGlnThrGlnLysArgArgLysAspTyrTyrMetProArgLeuAs 171
492 TAAATTTGACAGATGCAGACGCTGTGCA...ATGACAGAGTGACATGA 538
171 pLysPheValThrGluValAlaProIleGluLysSerThrAlaSerSera 188
539 GGGGGAATACCTATTCGATTAAGAAATAATCCGAGCGTGTCGCATC 588
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
589 GGCTCAGGACACCACTATTCGGCTTATGATGACAAACAGCGGATTT 638
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp.. 216
639 ATCCTATCCGGCGCATGGTAAATGGCGCATATACATATGACGGGT 688
217 .AsnTyrSer.....LeuIleLeuAsnAsnHisGluValGly. 228
689 GGGGAATATAT.....GGCGTANTT 708
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
709 AGTTTGAAGCGCGCATGTCGCCATGCCAACGAC.....TANCG 746
245 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheG 261
747 C.....C 748
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278
749 CTATGCCGATTCGACGTGCGGACGAGCGACAGCGGTTCGCAATGTTTAT 798
278 rIleLeuThrAsnTyrAlaValAlaLeuGlyAspSerGlySerProLeuPheVal 294
799 TATGACAAACAAACAATTAATGCTGCTCAACGAGTTTACAAACCGG 848
295 TyrAspArgGluLysGlyLysTrpLeuPheLeuGly.....Se 307
849 CTACCTTATTCGCGAGGAAACGTTTCCAGCTGATGACAAAGATT 898
307 rTyrAspPheTrpAlaGlyTyrAsn.....LysLysSer 319
899 GGTTCACGATGACATTTACAGAGCGCATACATACCGTCTNTTTTGA 948
319 rPlnGlnTyrPheAsnIleTyrLysProGlu..... 328
949 CCGCGCAGTACGCGACATTTTCCCTTACATCACACACAGCGTACG.. 996
329 .....PheAlaLysThrValIleuAspLysAspThrAl 339
997 .GGTAGGTAACAGAAACACGAAAGATTC...AATCCAAAGCTTA 1042
339 aGlySerLeuThrGlySerAsnThrGlnTyrAsnTrpAsnProThrGlyL 356
1043 AAGTACAGACAGTCCGACTGTTTGACGAATCTTTGAAT..... 1080
356 yThrSerValIleSerAsnGlySerGlnSerLeuAsnValAspLeuPhe 372
1081 .....GAACTGATTAAGAACACAGTTTACGCGGACGGGCTGT 1118
373 AspSerSerGlnAspThrAspSerLys..... 381
1119 TATTCAGTACCGTCCAAAGGTTAAACAGGTGAACCTTCTGTTTATCG 1168
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391
1169 ATTAGGCAACGGCAAACTCATCTTATCAACAACATCAACCAAGCGCG 1218
392 .ArgGlySerGlyThrLeuThrLeuAsnAsnHisLeuAspGlnGlyAla 407
1219 GGGGCTTGTATTTTGAAGGTGATTTTACGCTC...TCGGCTGAAACAA 1265
408 GlyGlyLeuPhePheGlnGlyAspTyrGluValLysGlyLysThrSerSp 424
1266 CGAACGCTGGCAACGCGCGGCTTCATATCATGAGACAGTACGCTTA 1315
424 rThrThrTrpLysGlyAlaGlyAlaSerValAlaAspGlyLysThrVal 441
1316 CTTGAAAGTAAACGGCGCTGGCAACGCGCGCTGTCCAAATTCGGCAA 1365
441 hrTrpLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
1366 GGCAGCGTGCACGCTTCAAGCCAAAGGGAAACACGCTCATGACGCT 1415
458 GlyThrLeuIleValGlnGlyLysGlyLysAsnLysGlySerLeuLysVa 474
1416 GGGCGACGATACAGTCATTTTGATCAGACGACAGATTAAGCAAAA 1465
474 IGLYAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysV 491
1466 AACAGCCTTTAGTGAATGCGCTTGTACGCGGACGAGGATAGGTGCA 1515
491 aLysAlaPheSerGlnValGlyIleValSerGlyArgSerThrValVal 507
1516 CTGAATGCCGATATACGTTTCAACCCGACAACTCATTTTCGCTTCG 1565
508 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheAr 524
1566 CGCGGACGTTTGATTTAAACGGGCACTGCTTCCTTCACCGATTC 1615
524 gGlyGlyArgLeuAspAlaAsnGlyLysAsnLeuThrPheGlnHisIleA 541
1616 AAAATPACCGATGAAGGCGCATGTTGNCNATCTATATGCCACACACA 1665
541 rGAsnIleAspAspGlyAlaArgLeuValAsnHisAsnTrpSerLysThr 557
1666 TTCACCGCTTACATTCACGAGGATGAAGTATATACACACCGAGT... 1710
558 SerThrValThrIleThrGlyLysLeuIleThrAspProAsnThrI 574
1710 ..... 1710
574 eThrProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheA 591
1710 ..... 1710
591 rGArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThr 607
1710 ..... 1710
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624
1711 .....GTAAGAAAT..... 1719
624 nSerGlyLysSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAsp 641
1720 .....ATCATAGACTTATTTACACCAAAAGAAATTCGCC 1752
641 lAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnGly 657
1753 TACAACGGTGTGTTGGGAGAGAAGATACGACCAAAACGAGCGCGCT 1802
658 PheAsnGlyTyrPheGlyGlnGlnGlu...GlyLysAsnAsnGlyAsnLe 673
1803 CAACCTGTTCACAGCCCGCGCAGAGACGACGACCCGCTGCTTCG 1852
673 uAsnValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuLeuThrG 690
```

```
1853 GCGGAACAATTTAAACGGCAACATCATCGCAACAAACGCAACGTTT 1902
|||||
690 LYLGLYHRSNLEUASNGLYASPLEUYSVALGLULYSGLYTHRIEUPHE 706
|||||
1903 TTCAGCGGACAGACCGACCGCATCAATCATTTAGAACGGGGTG 1952
|||||
707 LEUSERGILYRGPRTHTHRPHRHSALARGASPILEAGLYILESESE 723
|||||
1953 GTCAAAATGGAAGT.....ATCCCAACAGAGAAATCCGTGGGACA 1996
|||||
723 rthrlslysbpglnhlsphelaglulsnasngluvalValVal 740
|||||
1997 ACAGCTGATCAACCGCATGTTAAAGCGGAAATTCATATTCAGGCG 2046
|||||
740 spAsprlpleasntrgasnphenylsalathrasnllasnValthr 756
|||||
2047 GGGCAGCGCGTATTC...CGCAATGTTGCCAAAGTGGAAGCGGATG 2093
|||||
757 AsnAlatHrLeutyrserelYarYasnValAlaasnllthrserasnll 773
|||||
2094 NFATTTGACCAATCACGCCCAAGCAGTTTGTGTGCGACCGCATCAA 2143
|||||
773 ethrlaseraspsasnAlalysValHlslelYtr.....LysAlag 788
|||||
2144 GCCATACATCTGTACAGCTTGAGCTGACNGCTGTGCAAAATGTGTC 2193
|||||
788 lYAsprHrValCysValArgserAsprYtrHrGlyTrValthrcysHr 804
|||||
2194 GAANAANACATTAACGAGCATTAAGTATGTTGTTGACTAGACAGCA 2243
|||||
805 ThrAsprLysLeuSer...AsprYsalaleuasnSerPheasnAlatHrAs 820
|||||
2244 CNTNAGCGGCANTGTNAGNCTNCCNATNAGNTNNTNNAANCTCNCNG 2293
|||||
820 nValser..... 822
|||||
2294 GGCNTGCNNCACTNANAGCAATCTTAGCAATGGCATACACGTTAT 2343
|||||
822 ..... 822
|||||
2344 ACAGTCAGCCACAAACGCCAACCAACCTTAGCTGTGGGCA 2393
|||||
823 .....GlyasnValasnLeuSerGlyAs 830
|||||
2394 TGCCCAACACATTTAATCAAGCCACATTAAGCGCAACNCATCGGNTT 2443
|||||
830 nAlaasnPhelValleuGlyLysAlaasnLeuPhelGlyThrleSerGlyT 847
|||||
2444 CGGGCAATGCTTCAATTAATCAAGCAACACCGCGCAAAACGGCAGT 2493
|||||
847 hnglYasnSerelInValArgleutHrGluasn..... 857
|||||
2494 CTGACGCTTTCGACAAACGCTAAGCAACGTAAGCCATTCGCACTCAA 2543
|||||
857 ..... 857
|||||
2544 CGGCATGTCTCCCTAGCCGATTAAGCAGTATTCATTTGAAAAACAGCC 2593
|||||
857 ..... 857
|||||
2594 GCTTACCGCAACTCAGCGGACGCAAGANACAGCATTTACTTAAAA 2643
|||||
857 ..... 857
|||||
2644 GACAGCGAATGACGCTGCGCTCAGGACGCAATTAAGCAATTTAAACT 2693
|||||
858 ...serHstrprHlsleutHrGlyAsprserasnValasnGlnleuasnle 873
|||||
2694 TGACAGCGCAACATTAACATTCGCGCTATTCGCCACAGATCTCTCAG 2743
|||||
873 uAsprYsGLYHlsleHlsleuasnAla.....GlnasnAspAlaasnL 888
|||||
2744 GCGGCAAAACGCGCAGNGTGTACAGACGCGCGCGCGTTCGCGCGGT 2793
|||||
```

```
888 yValtrHrThr..... 891
|||||
2794 TCCCTATTATCCGTTACACCGCAACTTCGGTAGAATCCGTTTACAC 2843
|||||
892 .....TyrasnTh 894
|||||
2844 GCTGACGCTAAACGCAAAATTAACNGTCAAGAACATTCGCTTATGT 2893
|||||
894 rleutHrValasn...serleuSerGlyAsnGlySerPheYtrYleut 910
|||||
2894 CGGAACCTTCGCTACCGCAACCAATTTGAAGCTGGCGGAAAGTCC 2943
|||||
910 hrAspleuSerasnLysglInelYasnpyValValtrHrYsSerAla 926
|||||
2944 GAAGNACTTACACCTTGCGGCTCAACATACCGGCAACACCGGTAAG 2993
|||||
927 ThrGlyasnPherHrleuGlnValAlaAsprLysThrGlyLutProHrly 943
|||||
2994 CCTGATCAATTTGACGCTAGTGGAGGAAAGACAAACCGCTGTCCG 3043
|||||
943 s...AsnGlnleutHrleuPheAspAla.....SerasnAlatHrArg 957
|||||
3044 AAACCTTAATTTCAACCTGCAAAACGACACGCGATCGCGGCGGTG 3093
|||||
957 snAsnleuAsnValSerleuValGlyasnHrValAspleuGlyAlatP 973
|||||
3094 CGTTACCACTCATCCGCAAAAGACGCGGAGTTCGCTGCATATCCG.. 3141
|||||
974 LysTrLysleuHrArgasnValAsnGlyArgTyAspleuYrAsnProgl 990
|||||
3141 ..... 3141
|||||
990 uValGluLysArgasnGlnThrValAsprHrHrAsnllthrProa 1007
|||||
3141 ..... 3141
|||||
1007 snAsnllleGlnAlaAspValProSerValProSerAsnAsnGluIle 1023
|||||
3141 ..... 3141
|||||
1024 AlaArgValGluThrProValProProProAlaProAlaThrProSerG 1040
|||||
3141 ..... 3141
|||||
1040 uThrThrGluThrValAlaGluasnSerLysGlnGluSerLysThrValG 1057
|||||
3141 ..... 3141
|||||
1057 luyAsnGlnGluAspAlatHrGluThrThralaGlnasnGlyluVal 1073
|||||
3141 ..... 3141
|||||
1074 AlaGluGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGluVa 1090
|||||
3142 .....GTCAAAG 3148
|||||
1090 lAlaGlnSerGlySerGluThrGlnGluGluThrGlnThrGlnleLysG 1107
|||||
3149 AA..... 3150
|||||
1107 luthrAlaLysValGluLysGluGluLysAlaLysValGluLysAspGlu 1123
|||||
3151 .....CAAGGCTTTCGACAAACCTCGGCA 3176
|||||
1124 lIleGlnGluAlaProGlnMetAlaSerGluTrnSerProLysGlnAlaLys 1140
|||||
3177 G..... 3177
|||||
1140 sProAlaProLysGluValSerThrAspHrLysValGluGluThrGlnInv 1157
|||||
3177 ..... 3177
```

```

1157 aLglnAlaGlnProGlnInhrGlnSerThrThrValAlaAlaAlaGlnAla 1173
3178 .....GCAGACGCCAAAAACGGGGGAAAAAGA 3206
1174 ThrSerProAsnSerLysProAlaGlnGlnThrGlnProSerGlnLysThr 1190
3207 CAACGGCAAAAGCCCTGGACGGCGATGGCGCGG..... 3243
1190 rAsnAlaGlnProValThrProValValSerLysAsnGlnInhrGlnsnt 1207
3244 .....CGCATGCCCGCCGAAAAAGACAGAA 3267
1207 hrThrAspGlnProThrGlnArgLysThrAlaLysValGlnLysGln 1223
3268 AGCGTGGCCGAACCGCGCGCGGCGAGC..... 3297
1224 LysThrGlnGlnProProGlnValAlaSerGlnAlaSerProLysGlnGln 1240
3298 .....GGGAAATATGTC 3309
1240 GlnInhrGlnThrValGlnProGlnAlaValLeuGlnSerGlnLysValP 1257
3310 ..GGCATATATGAGCGGAGAGAGAAAAAGGGTCCAGCGGATATAA 3357
1257 roThrValAsnAsnAlaGlnGlnValGlnAlaGlnLeuGlnInhrGlnInhr 1273
3358 GAC..... 3360
1274 SerAlaThrValSerThrLysGlnProAlaProGlnAsnSerLysnThr 1290
3361 .....ACGCNTTGGCGAAACAGCGGAAACCGGAAACCGCGCGG 3400
1290 rGlySerAlaThrAlaIleThrGlnThrAlaGlnLysSerAspLysProG 1307
3401 nNACC..... 3405
1307 InThrGlnThrAlaAlaSerThrGlnAspAlaSerGlnHisLysAlaAsn 1323
3406 .....ACGCGCTTCCCGCG 3419
1324 ThrValAlaAspAsnSerValAlaAsnAsnSerGlnSerSerAspProly 1340
3420 CGCCCGCNCGCCCGCGGATTTGCCGACCGCAG..... 3456
1340 sSerArg...ArgArgArgSerLysSerGlnProGlnGlnInhrSerAlaG 1356
3456 ..... 3456
1356 LnglnThrThrAlaAlaSerThrAspGlnThrThrAlaAlaAspAsnSer 1372
3457 CCCCAACCGCAACCTCAACCCCAACCGCAG..... 3486
1373 LysArgSerLysProAsnArgArgSerArgSerValArgSerGlnLysP 1389
3487 .....CGCGACCTGA 3496
1389 oThrValThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuT 1406
3497 TNAGCGGTATGCCATAGCGGTGGATGATTTCCGC..... 3537
1406 hrSerThrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGln 1422
1423 rHeValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1439
3538 .....ACGCTCAACAGCGTTTCCGCTACAGAGAGAAATTTGACCGCGCT 3581
3582 G...TTTGGCGAAGACCGCGCAACGCGTTGGACAGACGCGCATCCGCGN 3628
1439 uGlnMetAsnAsnGlnGlyGlnThrAsnValThrValSerAsnThrSerM 1456
3629 ACACCAAAACACTACGCTTGGCAAGATTTCCGCGCTACGCGCAACAAAC 3678
1456 etAsnGlnsntYrSerSerSerGlnYrArgRgPheSerSerYrSer 1472

```

```

3679 GACCTGCGCCAAATCGGTATGCGAGAAAACTCGCAGCGG...CGCGT 3725
1473 ThrGlnThrGlnLeuGlnLysThrAspGlnThrIleSerAsnValGlnLe 1489
3726 CGGCATCTGCTGTTTCCACACCGGACCGGAAACAMCTTCGACGACGCA 3775
1489 uGlnGlyValPheThrGlyValArgAsnSerAsnAspLysAla 1506
3776 TCGGCAACTCGCGCGCTTGGCCACGCGCGCTTTTCGGCAATACGCGC 3825
1506 etSerLysAsnThrLeuAlaGlnValAsnPheGlySerLysGlyTyrAla 1522
3826 ATCGCGAGTTTGCACATCGGCATCGACACGCGCGCGGTTTACGACGCG 3875
1523 AspAsnHisThrPtyrLeuGlyLysAspLeuGlyTyrGlyLysPheGlnSe 1539
3876 CANTCTNTCAGACGCGATCGGAGGCAAAATCCGCGCGCGCTGCTGAT 3925
1539 rAsnLeuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnP 1556
3926 ACGGCATTCAGCGACGATACCGCGCGGTTTCCGCGATTCGCGATCGAA 3975
1556 heGlyLeuThrAlaGlyLysAlaPheAsnLeuLysnPheGlyIleThr 1572
3976 CCGTACATCGCGCGCACGCGCTATTTGTCGCAAAAACGGATTAACGCTA 4025
1573 ProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPheAla 1589
4026 CGAAACGTCATATGCGCCACCGCGGCTTGGCTTCACCGTACCNG 4075
1589 uAlaLysAspArgIleLysValAsnProIleSerValLysThrAlaPhe 1606
4076 CGGCATTAAGCAGATATTATTCATCAACCGCGCGCACACATATTCATC 4125
1606 IAGlnValAspLeuSerTyrThrThrHisLeuGly...PheSerVal 1621
4126 ACNCTTATTTNAGCTGTCTAT...ACCGATCGCGCTTGGGCAAGT 4172
1622 ThrProIleLeuSerAlaArgTyrAspThrAsnGlnLysSerLysI 1638
4173 CCGACACGCGGTCAATACCGCNGTATGCTCAGGATTCGGCAAAACC 4222
1638 e.....AsnValAsnGlnTyrAspPheAlaTyrAsnValGlnAsnGln 1653
4223 GCACTGCGCAATGCGCGGTAAACCGCGAAATCAAGGTTTCACGCTGTC 4272
1653 lnglnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1669
4273 NTCCAGCCTGCGCGCGCAAGNCCGACACTGGAAGCGCAACACGCGC 4322
1670 LeuIleGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1686
4323 GGGCATCAATTAAGCTAC 4341
1686 aGlnLeuLysLeuSerPhe 1692

seq_name: p1r2:A41859

seq_documentation_block:
19a-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C:Species: Haemophilus influenzae
A:Variety: Strain HK715
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: A41859
R:Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A:Reference number: A41859; M01D:92234949
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1702 <POU>

```

A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907
 A:Experimental source: strain HK715
 A>Note: sequence extracted from NCBI backbone (NCBIR:97282)
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1517.00 Length: 1881
 Ratio: 1.651 Gaps: 53
 Percent Similarity: 48.857 Percent Identity: 24.615

alignment_block:

US-09-303-518d-651 x A41859 ..

Align seg 1/1 to: A41859 from: 1 to: 1702

```

64 CGCTTCGCCGCTTACTTACCATATGCCGTGCTGGCATTCCTTC 113
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
5  LysPheLysLeuAsnPhelLeaLeuThrValAlaTyrAlaLeuThrPr 21
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
114 CCAGCTTGCGCGGACACACTTATTTGGCATCACTCACTCAATCTATC 163
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
21  OyrLthrGluAlaLeuValAlaArgAspAspValAspTyrGlnIlePheA 38
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
164 GCGACTTTCGCGAAATTAAGCAAGTTGCACTGGCGCGCAAGATATT 213
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
38  IArgPheLeaLeuAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
214 GAGGNTTCAACAAAAGGGAGTGTGCGCAATCAATGACAAAAGC 263
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
55  GluValAlaArgAspLysAsnAsnHisSerLeuGlyAsnValIleProAsnG 71
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
264 C...CCGATGATGATTTTCTGTGTGTCG...CGTAACGGCGTGGCG 307
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
71  YleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
308 CATTCGTGGCGCATATATATTTGACCGTGGCAGCAT...AACGGCGC 354
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
88  hLeuIleAsnProGlnTyrValAlaGlyValLysHisValSerAsnGly 104
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
355 TATAACAGCTGATTTTGGT.....GGCGAAGGAGAGAA 389
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
390 TCCCGATCAGACCGT.....TTTCTTACCAATGTG 421
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
121 nAspLysSerHisArgAspValSerSerGluAsnArgTyrPheSerV 138
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
422 TGAAGAAGATATATAT.....AAG 441
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
138 aIGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
442 CCTGACATTCACACCCCTTACACGGGATATACATATGCCGCTTGCA 491
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
155 GluAspGlnThrGlnLysAlaArgGluAspTyrTyrMetProAspGluAs 171
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
492 TAAATTTGTCACAGATCGAAGACCTGTGAA...ATGACGAGTGACATGA 538
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
171 PLYSPheValThrGluValAlaProIleGluAlaSerThrAlaSerSera 188
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
539 GGGGGAATACCTATTCGATTAACAAAATATCCGAGCGTGTCCGATC 588
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
188 sPalagLysThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
589 GGCTCAGGACACACTATTTGGCTTATGATGATGATGACAAACAGCGGATT 638
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp.. 216
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
639 ATCTTACTCCGGCGCATGTTAATTTGGCGGCAATATACATATGAGGGTT 688
   | ::::::::::: ::::::::::: ::::::::::: :::::::::::
217 AsnTyrSer.....LeuIleLeuAsnAsnHisGluValGly. 228

```

```

689 GGGGAATAT.....GGCTANTT 708
   | ::::::::::: :::::::::::
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
   | ::::::::::: :::::::::::
709 AGTTTGAGCGCGGATGTGCCCATGCCAAGC.....TATGG 746
   | ::::::::::: :::::::::::
245 GlyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheCl 261
   | ::::::::::: :::::::::::
747 C.....C 748
   | ::::::::::: :::::::::::
261 yAsnSerLysGluLuhHisSerAspProLysGlyIleLeuSerGlnAsp 278
   | ::::::::::: :::::::::::
749 CTATGCCGATTTGCGGTGGCGAGCGGCGGTTGGCAATGTTATT 798
   | ::::::::::: :::::::::::
278 rLeuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheVal 294
   | ::::::::::: :::::::::::
799 TATGACAAACAAACAAATTAATGGCTGTCACAGAGTTTACAAACCG 848
   | ::::::::::: :::::::::::
295 TyrAspArgGluLysGlyLysTyrPhePheLeuGly.....Se 307
   | ::::::::::: :::::::::::
849 CTACCTTATTCGCGAGGAAACGGTTTCCAGCTGATACGCAAGATT 898
   | ::::::::::: :::::::::::
307 rTyrAspPheThrPalagLysTyrAsn.....LysLysSerT 319
   | ::::::::::: :::::::::::
899 GGTCTACGATGACATTTTACAGAGCGATACATACCTCTTTTGA 948
   | ::::::::::: :::::::::::
319 rPglngLuhTyrAsnIleTyrLysProGlu..... 328
   | ::::::::::: :::::::::::
949 CCGCGCAGTAAACGACATTTTCTTACATCCACAAACAGCGTACG.. 996
   | ::::::::::: :::::::::::
329 .....PheAlaLysThrValLeuAspLysAspThrAl 339
   | ::::::::::: :::::::::::
997 ..GGTACGGTAAACAAACAAACGAAAGTNTTC...AATCCAAACCTTA 1042
   | ::::::::::: :::::::::::
339 aGlySerLeuThrGlySerAsnThrGlnTyrAsnThrAsnProThrGlyL 356
   | ::::::::::: :::::::::::
1043 AAGTACAGACAGTCCGACTGTTGACGAATCTTGAAT..... 1080
   | ::::::::::: :::::::::::
356 yStrSerValIleSerAsnGlySerGluSerLeuAsnValAspLeuPhe 372
   | ::::::::::: :::::::::::
1081 .....GAACTGATTAAGAACGATTTACGCGGAGGGGTGT 1118
   | ::::::::::: :::::::::::
373 AspSerSerGlnAspThrAspSerLys..... 381
   | ::::::::::: :::::::::::
1119 TAATCAGTACCCGCAAGCTTAACAAACGCTGAACCTTTCTTATATG 1168
   | ::::::::::: :::::::::::
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391
   | ::::::::::: :::::::::::
1169 ATTACGCAACGCGCAAACTCATCTATCAACAAACATCAACGAGCGG 1218
   | ::::::::::: :::::::::::
392 ..ArgGlySerGlyThrLeuThrLeuAsnAsnHisLeuAspGlnGlyAla 407
   | ::::::::::: :::::::::::
1219 GCGGTTTGTATTTGAAGGTGATTTACGGTC...TCGCGTGAACAAA 1265
   | ::::::::::: :::::::::::
408 GlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThrSerAsp 424
   | ::::::::::: :::::::::::
1266 CGAAAGTGGCAAGCGCGGCTCATATACAGTAAAGACAGTACCGTTA 1315
   | ::::::::::: :::::::::::
424 rThrThrThrLysGlyAlaGlyAlaSerValAlaAspGlyLysThrValT 441
   | ::::::::::: :::::::::::
1316 CTGGAAGATTAACGCGGTGGCAACGAGCGCTGCAAAATGGGCAAA 1365
   | ::::::::::: :::::::::::
441 hTrThrPysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
   | ::::::::::: :::::::::::
1366 GGCAGCTGCAGCTTAACGCAAAAGGGAAACCAAGGCTGCATGACGCT 1415
   | ::::::::::: :::::::::::
458 GlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySerLeuLysVa 474
   | ::::::::::: :::::::::::
1416 GGGCGAGGTACAGCATTTTGATGATCAGAGGACAGCATTAAGCAAAA 1465
   | ::::::::::: :::::::::::
474 IGLysAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysV 491
   | ::::::::::: :::::::::::
1466 AACAAACCTTTAGTGAATCGGCTTGNTCAGCGGAGGGGTACGTCGAA 1515

```

```

491 aLLysAlaPheSerGlnValGlyLeuValSerGlyArgSerThrValVal 507
      :::::::::::::::::::: ||::::::::::::::::::
1516 CTGAATGCCGTAATCATGTTCAACCCGACAACTCATTTGGCTTCG 1565
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
508 LeuSnaAspArgLysGlnValAspProAsnSerIleTyrPheGlyPheAr 524
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1566 CGGCGAGCGTTTGGATTAAACGGCATTCGCTTCGTCACCGTATTC 1615
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
524 gGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIleA 541
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1616 AAAATACCGATGAAGGCGCATGATGTCNATCATTAATGCACAAACA 1665
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
541 rGAsnHisLeuAspArgLysAlaArgLeuValAsnHisAsnThrSerLysThr 557
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1666 TCCACCGTTACCATTCACAGGAATGAAGTATTCACACCGAGT.... 1710
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
558 SerThrValThrIleThrGlyLeuSerLeuIleThrAspProAsnThrI 574
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1710 ..... 1710
574 eThrProTyrAsnHisLeuAspAlaProAspGluAsnProTyrAlaPhea 591
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1710 ..... 1710
591 rGArgLLeuLysAspArgLysGlnLeuTyrLeuAsnLeuGluAsnTyrThr 607
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1710 ..... 1710
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1711 ..... 1711
624 nSerGlyLeuSerAsnGluAsnThrProLeuTyrMetGlyLysThrSerAsp 641
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1720 ..... 1720
641 LuAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnLys 657
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1753 TGCAGAGCGTTGGTTGGGAGAGAAAGATACGCCAAAGCAAGGCGGCT 1802
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
658 PheAsnGlyTyrPheGlyLeuGluGlu... GlyLysAsnAsnGlyAsnLe 673
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1803 CAACCTGTGTTACAGCCCGCGGAGAAAGCCGACCCGCTGCTTCG 1852
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
673 uAsnValThrPheLysGlyLysSerGluGluAsnArgPheLeuLeuThr 690
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1853 GGGCAACAATTAAACGCAACATCAGCAAAACGCAAACTGTT 1902
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
690 LyGlyThrAsnLeuAsnGlyAsnLeuLysValGlyLysGlyThrLeuPhe 706
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1903 TTCAGGCGAGAGACGACACCGACGCTACATCATTTAGCAAGCGGGTG 1952
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
707 LeuSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSe 723
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1953 GTCAAAATGAGAGT.... ATCCCAACAGAGAAATCGTGGGACA 1996
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
723 rThrLysLysAspArgLHisPheAlaGluAsnAsnGluValValGluVal 740
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1997 AGCAGCTGATTCNAACCGCATTTAAAGGAAATTTCCATATTACAGGC 2046
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
740 sPAspTrpLLeuAsnArgAsnPheLysAlaThrAsnHisLeuValThrSna 756
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2047 GGGCAGGCGGTGATTTCC... CGCAATGTTGCCAAAGTGGAAGCGCATG 2093
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnI 773
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2094 NCATTTGACCAATCACGCCCAAGCAAGTTTGTGTGCGCACCGCATCAA 2143
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
773 eThrAlaSerAspAsnAlaLysValHisIleGlyTyr.... LysAlaG 788
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2144 GCATTCATCTGTACAGCTGGAGCTGACGNGCTCGCAAAATGTGTC 2193
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||

```

```

788 LysPThrValCysValArgSerAspTyrThrGlyTyrValThrCysThr 804
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2194 GAANAANCAATTACGACGATTAAGTATGCTTCATTGACATTAACGNA 2243
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
805 ThrAspLysLeuSer... AspLysAlaLeuAsnSerPheAsnAlaThrAs 820
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2244 CNTNACCGGCAGTGTAGNCTNCCNATNAGCNTNTNTTNAANCTCNCNG 2293
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
820 nValSer..... 822
2294 GGCNTGCTNCCACTNAANGCAATCTTAGTCAAAATGGCGATACAGTTAT 2343
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
822 ..... 822
2344 ACAGTCAGGCACACGCCAACCCAAACGCAACCTTAGCGTGGGCAA 2393
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
823 ..... GlyAsnValAsnLeuSerGlyLys 830
2394 TGCCCAAGCAACATTTAATCAACCCACATTTAAACGGCAACNCATCGGNTT 2443
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
830 nAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847
2444 CGGCGCAATGCTTCAATTATCTAAGCAACACGCCCGCACAAAACGCGAGT 2493
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
847 hrGlyAsnSerGlnValArgLeuThrGluAsn..... 857
2494 CTGACGCTTCCGCAACGCTAAGCAACGTAAGCATTCGCACTCAA 2543
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
857 ..... 857
2544 CGGCAATGCTCTCCCTAGCCGATAGCAAGTATTCATTGAAACAGCC 2593
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
857 ..... 857
2594 GCCTTACCGGCAACACTCAGCGGCGACAGAGNACAGCATTAACCTTAA 2643
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
857 ..... 857
2644 GACAGCGAATGACGCTCCGCTGAGCGACGGAATTAAGCAATTTAACT 2693
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
858 ...SerHisTrpHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnLe 873
2694 TGACACGCCACATTAACACTCAATTCGCGCTTCCGCAACGATGCGCAG 2743
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
873 uAspLysGlyHisIleHisLeuAsnAla.... GlnAsnAspAlaAsnL 888
2744 CGCGCAAAACCGGACGNGTGTACAGACGCGCGCGCGCTTGCGCGCGT 2793
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
888 yValAlThrThr..... 891
2794 TCCCTATTATCCGTTACACGCCAATTCGCTGCAATCCGTTCAACAC 2843
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
892 ..... TyrAsnTh 894
2844 CCTACGCTTAACGCAATTTGAACNGTCAAGGACATTCGCTTATGT 2893
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
894 rLeuThrValAsn... SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910
910 hrAspLeuSerAsnLysGlnGlyAspLysValValValThrLysSerAla 926
2944 GAAAGNACTTAACCTTGGCGGCTCAACAATACGGGCAACACCGGTAAG 2993
      ||||| ::||| ::||| ::||| ::||| ::||| ::|||
927 ThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyLeuProThrLy 943
2994 CCTGATCAATTAAGCTAGTAGGAGGAAGAAACAAACACCGCTGCGC 3043
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
943 s... AsnGlnLeuThrLeuPheAspAla.... SerAsnAlaThrArgA 957
3044 AAAACCTTAATTCACCTCGCAAAACGACACGTCGATGCGGCGCGCTGG 3093
      ::||| ::||| ::||| ::||| ::||| ::||| ::|||
957 sAsnLeuAsnValSerLeuValGlyLysnThrValAspLeuGlyAlaTrp 973

```


3094 CGTTACCACTCATCCGCAAGAGCGGAGTTCCGCTGCATATACG... 3141
:::|||||::: |||:::|||||
974 LysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyrAsnProG1 990
3141 3141
990 uValGluLysArgAsnGlnThrValAspThrThrAsnIleThrThrProA 1007
3141 3141
1007 snAsnIleGlnAlaAspValProSerValProSerAsnAsnGluIle 1023
3141 3141
1024 AlaArgValGluThrProValProProAlaProAlaThrProSerG1 1040
3141 3141
1040 uThrThrGluThrValAlaGluAsnSerLysGlnLysThrValG 1057
3141 3141
1057 LuLysAsnGluGlnAspAlaThrGluThrAlaGlnAsnGlyGluVal 1073
3141 3141
1074 AlaGluGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGluVal 1090
3142 GTCAAG 3148
1090 lAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGluIleLysG 1107
3149 AACAGAGCTTCCGACAACTCCGCAAGCGAGAGCCCAAAACAG... 3195
|||:::|||||:::|||||:::|||||
1107 LuThrAlaLysValGluLysGluGluLysAlaLysValGluLysGlu 1123
3195 3195
1124 LysAlaLysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSe 1140
3195 3195
1140 rGluThrSerProLysGlnAlaLysProAlaProLysGluValSerThrA 1157
3195 3195
1157 sPThrLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSer 1173
3195 3195
1174 ThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLysProAlaG1 1190
3196 GCGGAAAAAGACACAGCGCAAGCCTTACCGCGCTGA 3232
1190 uGluThrGlnProSerGluLysThrAsnAlaGluProValThrProVal 1207
3233 TTGCGCGCGGG... 3243
1207 alSerLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGlu 1223
3244 CGCGATGCCCGCCGAAAGACAGAAAGCGTCCGACCGCGCGCGCGC 3293
:::|||||:::|||||:::|||||
1224 LysThrAlaLysValGluThrGluLysThrGlnLysProGlnValAl 1240
3294 AGG... 3297
1240 aSerGlnAlaSerProLysGlnGluGlnSerGluThrValGlnProGlnA 1257
3298 GGGGAAATGTC...GCCATTATGACGCGAGAGAGAG 3333
:::|||||:::|||||:::|||||
1257 laValLeuGluSerGluAsnValProThrValAsnAsnAlaGluGluVal 1273

3334 AAAAACGGGTGACGGGATTAAGAC... 3360
:::|||||:::|||||:::|||||
1274 GlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnPr 1290
3361 AGCCNTTGGCGAAAC 3376
1290 oAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIleThrGluT 1307
3377 AGCGGAAAGCGGAACCGCGCGGNTAC... 3405
|||:::|||||
1307 hrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaSerThrGlu 1323
3405 3405
1324 AspAlaSerGlnIleLysAlaAsnThrValAlaAlaAspAsnSerValAlaAs 1340
3406 ACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGC 3445
:::|||||:::|||||:::|||||
1340 AsnSerGluSerSerGluProLysSerArg...ArgArgArgSerLies 1356
3446 CGCAACCGCAG... 3456
|||:::|||||
1356 erGlnProGlnGluThrSerAlaGluGluThrAlaAlaSerThrAsp 1372
3457 CCCCAACCGGAACCTCAACCCCAAC 3482
1373 GluThrThrIleAlaAspAsnSerLysArgSerLysProAsnArgSe 1389
3483 GCAG... 3486
1389 rArgArgSerValArgSerGluProThrValThrAsnGlySerAspArgS 1406
3487 CCGGACCTGATNAGCCGTTATGCCAATACCGCTTG 3522
|||:::|||||:::|||||
1406 erThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnAlaValIle 1422
3523 AGTGAATTTCCGCC... ACCGTCAACAGCGCTTTCCG 3557
|||:::|||||
1423 SerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLysAl 1439
3558 CGTACAGAGCAATTTGGACCGCGTG...TTTGGCAAGACCGCGCAACG 3604
|||:::|||||:::|||||
1439 aValSerGlnIleIleSerGlnLeuGluMetAsnGluGluGlyLysAla 1456
3605 CNGTTTGACACAGCNGCATCCGGAACACCAACACTACCTTGGCAAGAT 3654
|||:::|||||
1456 snValThrValSerAsnThrSerMetAsnGluAsnThrSerSerSerGln 1472
3655 TTCCGCGCTACCGCCCAACACCGACCTGCGCAATGCGTTCGAGA 3704
|||:::|||||
1473 TyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTyrAspG1 1489
3705 AAACCTCGGACGCGG...CGCGTGGCATCCTGTTTGGCACAMCGGA 3751
|||:::|||||
1489 nThrIleSerAsnAsnValGlnLeuGlyValPheThrTyrValArgA 1506
3752 CCGAAAAACNCTTCGACAGACGCGCATCGGCACTCGGACGGCTGCCAC 3801
:::|||||
1506 snSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaGlnVal 1522
3802 GCGCGCGTTTTCGGCAATACGCGATCGGACGATGCGATGCGCATCAG 3851
:::|||||
1523 AsnPheTyrSerLysTyrTyrAlaAspAsnHisThrTyrLeuGlyLysAs 1539
3852 CACGGGCGGCGGTTTTCGACGCGGCACTTCACAGCGGATCGAGGCA 3901
:::|||||
1539 pleuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHisAsnAlaL 1556
3902 AAATCGCGCGCGCGGTCGATACGCGATTCAGGATTCAGGATTCAGGCGCG 3951
|||:::|||||
1556 yPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyLysAlaPhe 1572
3952 GGTTCGGCGGATTCGCGATCGAACCGTACATCGCGGCAACCGCTATT 4001

762	GTTCSSAAAGTGTATTATGACAAACAAACAAATAAAGTGGCTGCAAC	831
279	lysertholeuphialaPhaasprylsdlilyusasnrlntprvalPhleu	295
832	GGAAGTTTACAAACGGGCTACCCATTATCCGGGACGAAACGGTTTCA	881
296	glYthrtYasprYtYtPrIaaglytYnGly.....	305
882	GCTGATACGCAAGATGGTTGTTACGATGACATTATACAGGCGAT...	927
306lyslYsertrpInglntprYasnrltYtYlYslYslgPha	320
928ACCAATACGCTGCTNTTTGAAACCGGCGAGTAAACGA	963
320	laasprYsllelyslGlnlIsaasrlnlaaglytYhvalYlYslslYnlll	336
964CATTTTCTCTTTACATCCAAACAAACAGGATGAGTACG	1007
337	glunlshlstrpYstYhrtYhrtglYtYhrtYasnserlnlleglySerthr...	352
1008	AGAAACCAAGAAAGGTNTCCAAATCCAAACGTTAAAGTACAGACATCC	1057
353	1107
1058	GACTGTTTGACGAATCTTGAATGAACGTATAAAGAACACAGTTTACCG	1107
355	rgleuallasn.....AsnGluglyAsp.....	362
1108	GCAGGGGTGTATATACGTACCGTCCAGGTTAAACAACGCTGAAACCT	1157
363AlaasnaInglYlnIsnIva	369
1158	TTCSTTTATGATTAAGGACACGCGCAACAACTATCTTATCAAAACATCA	1207
369	lYhrtPheluaSp.....AsnclYthrtleuvalleuasnGlnsnlIlea	384
1208	ACCAAGCGCGCGGCGTTTGATTTTGAAGGTGATTTTACGGTCTCGCT	1257
384	snGlnlelylaaglyslYleuPherPhelYslYasprYtYhrtYalYslY	400
1258	GAACAACAAGAA...ACGGGCAAGCGCGCGCGCTTCAATACAGTGAAGA	1304
401	AlaasnaasnaPrlethrtYhrtleuclYlaaglytYleasrvalAlaasgl	417
1305	CAGTACCGTTACTGGAAAGTAAACGGCGGTGGCAACGACCGCTGTCA	1354
417	YlYslYsValIaItrpGlnIvalYlYasPrIasnProasnolYasprYleuall	434
1355	AAATCGGCAAAAGCAACGCGGACGCTTCAACGCCAAAGGGAAACCAAGC	1404
434	YsllelelyslYstYhrtYhrtleuGlnleasnGlytYhGlyvalaIasnGlnly	450
1405	TGCAATACGCGTGGGCGACGCTACAGTCAATTTTGGATCAGCAGGACAGA	1454
451	GlnleuYsValaIGlyasprYlYhrtYalIleuasnGlnlYslalaaPral	467
1455	TAAAGCAAAAAACAACGCTTATAGTGAATGCGCTTGTCACGGCGCAGG	1504
467	asprYslYsValaGlnalIaPheserGlnvalaIGlyIlevalserlYatrg	484
1505	GTACGGTGCACGTGAATGCGATATACGTTCACACCCGACCAACCTGAT	1554
484	lYtYhrtleuvalleuasnserIasnGlnIleasnProasrlnleuYt	500
1555	TTCGGCTTTCGGGCGGACGTTTGGATTTAAACGGGCAATTCGCTTGGT	1604
501	PhelgYlYPhaIrgGlyGlyIaYtrleuAsrlnlaasnGlylYasnAsprleuYhPh	517
1605	CAACGGTATTCAAATACCGATGAAGGGGCGCATGATTTGNCNATCATATG	1654
517	eglnlshlstrpYasnValaasprYlYlaIaYtYlIlevalaIasnIasnrt	534

[illegible]

```
794 ..... AspLysAlaLeuAsnSerPheAspAlaThrArgIleAsn 806
2371 GGCAACCTTACGCTCTGTGGCAATGCCCAAGCAATTTAATCAAGCCAC 2420
807 GlyAsnValAsnLeuAsnGlnAsnAlaIleValLeuGlyLysAlaI 823
2421 ATTAAAGCGCAACATCGGNTTGGGCAATGCTTCATTTAATCTAAGCA 2470
823 AleuTrpGlyLysIleGlnGlnGlnLysSerArgValSerLeuAsnG 840
2471 ACAAGCGCGCAAAACGGCAGTCTGACGCTTCCGACAACGTTAAGCA 2520
840 LnhIS..... 841
2521 AACGTAAACCATTCGGCACTCAACGGCAATGTCCTAGCCGATAAAGC 2570
841 ..... 841
2571 AGTATTCATTTTGAACACGCCCTTTACCGGACAACTCAGCGGACGA 2620
841 ..... 841
2621 AGAAGAACAGCATTAACATTAAGACAGCAAGATGAGCGTCGCGTACGC 2670
842 ..... SerLysTrpHisLeuThrGlyAsp 849
2671 ACGGAATTAGGCAATTTAAACCTTGACACAGCCACCATTTACATCAATTC 2720
850 SerGlnValHisAsnLeuSerIleValAspSerHisIleHisLeuAsnAs 866
2721 CGCCATATGCGCACAGTCTGCAGCGCGCAAAACGGGACGNGTGTACAGCA 2770
866 nAla..... SerAspAlaGlnSerAla..... 873
2771 CGCCGCGCGCGCTTCGCGCGCTTCCTATTATCCGTTACACCGCAACT 2820
873 ..... 873
2821 TCGGTAGAAATCCCGTTTCAACACGCTGACGGTAACGGCAATTTGAACNG 2870
874 ..... AsnLysTyrHisThrIleLysIleAsn... HisLeuSerG1 886
2871 TCAAGGAACATTCGCGTTTATGTGCGAAGTCTTGGTACCGAAGCGACA 2920
886 yAsnGlnHisPheHisTyrLeuThrAspLeuAlaLysAsnLeuGlyAspL 903
2921 AATTGAAGCTGGCGAAGATTCCGAAGNACTTACACCTTGGCGGTCAAC 2970
903 yValLeuValLysGlnSerAlaSerGlyHisTyrGlnLeuHisValGln 919
2971 AATTACCGGCAACGAAACC..... GTAAGCTCGATCAATTGACGGTACT 3014
920 AsnLysThrGlyLupProAsnGlnGlnLeuAspLeuPheAspAlaSe 936
3015 GGAAGGAAGAGACACAAACCGCTGTCCGAACCACTTAATTTCACCCCTGC 3064
936 rSerValGlnAspArg..... SerArgLeuPheValSerLeu 949
3065 AAAAGCAACAGCTCGATCGCGCGCGGTGCGTTACCAACTCTCCGCAAA 3114
949 lAsnHisTyrValAspLeuGlyAlaLeuArgTyrThrIleLysThrGlu 965
3115 GACGCGGAGTTCGCGCTGCATTAAT..... CCGGT 3143
966 AsnGlyIleThrArgLeuTyrAsnProTyrAlaGlyAsnGlyArgPro 982
3144 CAAGAAGACAAGAGCTTTCGACAACCTCGCAAGGAGGAGCAAGCCAAAC 3193
982 lLysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA 999
3194 AGCGGAAAAAGACAACGCG..... 3213
999 lathrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValVal 1015
3214 ..... CAAAGCCCTTGACGGCTGATGTCGCGC 3239
1016 AlaProProSerProGlnAlaAsnGlnIleGlnIleAlaLeuArgGlnI 1032
3240 CGGCGCGGATGCGCGCGCAAAAGACAGAAAGCGTTGCCGACGCGCGCGC 3289
1032 nAlaLysAlaGlnGlnIleValLysArgGlnGlnIleAlaGlnIleGlnI 1049
3290 NCGCAGCGCGGGAATAATGTCGGCATTTATGACGCGGAGGAGCAAGAAAA 3339
1049 aAla..... ArgGlnLysAspGlnIleLys 1058
3340 CGGCTGCAGCGGAT..... 3354
1059 ArgLysAlaIleGlnIleAlaArgGlnGlnIleGlnIleAlaArgLysAlaI 1075
3355 ..... AAGACAGCGCNT 3367
1075 aGluLeuAlaIleLysGlnLysAlaGlnIleGlnIleGlnIleGlnIle 1092
3368 TGGCGAAACAGCGCGCAAGCGGAAACCGCGGNTTACACGCGCTTCCGC 3417
1092 euAlaArgGlnLysAlaGlnIleGlnIleSerHisGlnAlaAsnAlaLysPro 1108
3418 CGGCGCGCGNCGCGCGCGGATTTGCCGCAACCG..... 3453
1109 LysArgArgArgArgArgAlaIleLeuProArgProProAlaProValPh 1125
3453 ..... 3453
1125 eSerLeuAspAspTyrAspAlaLysAspAsnSerGlnSerIleGlyA 1142
3453 ..... 3453
1142 snLeuAlaArgValIleProArgMetGlyArgGluLeuIleAsnAspTyr 1158
3453 ..... 3453
1159 GluGlnIleProLeuGlnIleGlnIleAspGlnAlaGlnIleGlnIleArg 1175
3453 ..... 3453
1175 gGlnAlaThrGlnPheHisSerLysSerArgAsnArgAlaIleSers 1192
3453 ..... 3453
1192 eArgLupProSerSerAspGlnAspAlaSerGlnSerValSerThrSerAsp 1208
3453 ..... 3453
1209 LysHisProGlnAspAsnThrGlnLeuHisGlnLysValGlnIleThrAlaG1 1225
3454 ..... CAGCCC..... CAAACCGCAACCTCAACCCCAACCGCAGCGG 3490
1225 yLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAlaAlaAlaAla 1242
3491 ACCTGATNAGCGGTTATGCCAATAGCGGTTTGAATTTTCCGCGACG 3540
1242 spAlaValSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSer 1258
3541 CTCACAGCGTT..... TTGCGCGTACAGACGCAAT 3572
1259 ThrGlnSerIleLeuLeuAspThrGlnGlyAlaTyrLeuThrArgHisIleAl 1275
3573 GGACCGCGGTGTTGCCAAGACCGCGCAACGGGTTTGACAGACNGCA 3622
1275 aGlnLysSerArgAlaAspAlaGlnLysAsnSerValTyrMetSerAsn 1292
3623 TCCGNAACCAACACTACCGGTTGCGCAAGATTTCGCGGCTACCGGCAA 3672
1292 hrGlyTyrGlyArgAspTyrAlaSerAlaGlnIleGlnIleGlnIleGlnIle 1308
```

```

3673 CAACCGGACCTGGCCAAATCGGTATGACAGAAAACCTCGCGACCGGG.. 3720
    :: |||||.....:|||||
1309 LysArgThrGlnThrGlnIleGlyLeuAspArgSerLeuSerGluAsnMe 1325
    :|||:|||||.....:|||||
3721 ..CGCGTGGGACCTCTGTTTTCGACAAACCGGACCGAAACACNCCTGCAGC 3769
    :|||:|||||.....:|||||
1325 tGlnIleGlyGlyValLeuThrTyrSerAspSerGlnHisThrPheAspG 1342
    :|||:|||||.....:|||||
3770 ACGGCATCGGCAACCTCGGACGCGCTTGGCCACGCGCGCGTTTCGGGCA 3819
    :|||:|||||.....:|||||
1342 InAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeuTyrGlyLys 1357
    :|||:|||||.....:|||||
3820 TACGGCATCGGC...AGGTCGACATCGGCATCGGACGAAATCGCGCGGTTT 3866
    :|||:|||||.....:|||||
1358 TyrTyrLeuAsnAspAlaTyrPyrValAlaGlyAspIleLysAlaGly 1374
    :|||:|||||.....:|||||
3867 TACGACGCGCAATCTTTCAGACGGCATCGGAGCGAAATCGCGCGCGG 3916
    :|||:|||||.....:|||||
1374 rLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgThrS 1391
    :|||:|||||.....:|||||
3917 TGCTGATTCAGGCAATTCAGGCGCATACCGCGCGCTTCGCGGATTC 3966
    :|||:|||||.....:|||||
1391 erIleGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIleAsnGlnPhe 1407
    :|||:|||||.....:|||||
3967 GGCATCGAACCGTACATCGGCGGACCGCGCTATTGTCGCAAAAGCGGA 4016
    :|||:|||||.....:|||||
1408 GluIleValProSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAs 1424
    :|||:|||||.....:|||||
4017 TTCCCGCTCAGAAACGTCATATCCGACCGCGCGCTTCGCGTTCAC 4066
    :|||:|||||.....:|||||
1424 pTyrLysLeuGlyAspAspSerValLysValSerSerMetAlaValLysT 1441
    :|||:|||||.....:|||||
4067 GNTACCGGCGGCGCATTAAGCGAGATTATTCATTAACCGGCGCAACAC 4116
    :|||:|||||.....:|||||
1441 hrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsn 1456
    :|||:|||||.....:|||||
4117 ATATCCATCAACNCCTTATTNAGCCTGCTCTATACCGATCGCGCTTC 4166
    :|||:|||||.....:|||||
1457 LeuThrValLysProLeuLeuSerAlaAlaTyrPhe...AlaAsnTyrcl 1472
    :|||:|||||.....:|||||
4167 CAAGTCGCAACGCGGCTCAATACCGCNGTATGCTGACGATTCGCGCA 4216
    :|||:|||||.....:|||||
1472 yLysGlyGlyValAsnValGlyGlyLysSerPheAlaTyrLysAlaAsp 1489
    :|||:|||||.....:|||||
4217 AAACCGGACGTGCGGATGCGGCTAAACGCGGAAATCAAGCTTCACG 4266
    :|||:|||||.....:|||||
1489 snGlnGlnIleTyrSerAlaGlyValAlaLeuLeuTyrArgAsnValThr 1505
    :|||:|||||.....:|||||
4267 CGTCTCCTCAGCGTGGCGCGCGCAAGGCGCAACTGGAGCGCAACA 4316
    :|||:|||||.....:|||||
1506 LeuAsnValAlaAsnGlySerIleThrLysGlyLysGlnLeuGlyGln 1522
    :|||:|||||.....:|||||
4317 CAGCGGCGGCAATTAAGCTACCGCTCGG 4347
    :|||:|||||.....:|||||
1522 sSerGlyGlnIleLysIleGlnIleArgPhe 1532

```

seq_name: p1r2:C41859

seq_documentation_block:
 IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
 C:Species: Haemophilus influenzae
 A:Variety: strain HK613
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
 C:Accession: C41859
 R:Poulsen, K.; Reinholdt, J.; Kilian, M.
 J. Bacteriol. 174, 2913-2921, 1992
 A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
 A:Reference number: A41859; M01D:92234949
 A:Accession: C41859
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613
 A:Note: sequence extracted from NCBI backbone (NCBI:97285)
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

alignment_scores:
 Quality: 1475.00 Length: 1924
 Ratio: 1.517 Gaps: 64
 Percent Similarity: 50.520 Percent Identity: 25.364

alignment_block:

US-09-303-518d-651 x C41859 ..

Align seg 1/1 to: C41859 from: 1 to: 1849

```

64 CGCTTCGCGCGCTTACTTACGATATGCTGCTGCTTCGCTTCCTTC 113
    :|||:|||||.....:|||||
5 LysPheLysLeuAsnPheIleLeuThrValAlaTyrAlaLeuThrPr 21
    :|||:|||||.....:|||||
114 CCAAGCTTGGCGGACACACTTATTTGCGCATCACTACCAATACATATC 163
    :|||:|||||.....:|||||
21 cyrThrGlnAlaAlaLeuValArgAspAspValAspTyrGlnIlePhe 38
    :|||:|||||.....:|||||
164 GCGACTTTGGCGGAAATAAAGCAAGTTTGCAGTCGGGCGGCAAGATAT 213
    :|||:|||||.....:|||||
38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
    :|||:|||||.....:|||||
214 GAGGTNTCAACAAAAAGGAGTGGTGGCAATCAATCAATCAACAAAGC 263
    :|||:|||||.....:|||||
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
    :|||:|||||.....:|||||
264 C...CGATGATGATATTTTCTGTGCTG...CGTAACGCGCTGGCG 307
    :|||:|||||.....:|||||
71 yLeuPheMetIleAspPheSerValAlaAspValAspLysArgIleAla 88
    :|||:|||||.....:|||||
308 CATGCTGGGCGGATCAATATTTGTGACCGCTGCACAT...AACGGCGC 354
    :|||:|||||.....:|||||
88 hrLeuValAsnProGlnTyrValAlaGlyValLysHisValSerAsnGly 104
    :|||:|||||.....:|||||
355 TATAACCAAGTTGATTTTGGT.....CGGGAAGGAAGNAA 389
    :|||:|||||.....:|||||
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 121
    :|||:|||||.....:|||||
390 TCCCGATCAGACCGT.....TTTCTTACCAATTG 421
    :|||:|||||.....:|||||
121 nAlaLysSerHisArgAspValSerSerGlnGluAsnArgTyrThrV 138
    :|||:|||||.....:|||||
422 TGAAGCAATATATTATAGCTGACAT..... 450
    :|||:|||||.....:|||||
138 alGluLysAsnAsnPheProThrGlnAsnValThrSerPheThrLys 154
    :|||:|||||.....:|||||
451 .....TCACACCTTACACGCGGATTTANCAATTCGCGGCTTT 488
    :|||:|||||.....:|||||
155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArg 171
    :|||:|||||.....:|||||
489 GCATAAATTTGTCACAGATGACAGAACTGTGCA...ATGACGAGTGACA 535
    :|||:|||||.....:|||||
171 uAspLysPheValThrGluValAlaProIleGlnAlaSerThrAlaAsn 188
    :|||:|||||.....:|||||
536 TGAAGGGGAATACCTATTCGATTAAGAAATAATCCGCGGCTGCGCG 585
    :|||:|||||.....:|||||
188 snAsnLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArg 204
    :|||:|||||.....:|||||
586 ATGCGCTCAGGACACCATAT.....TGCGGTAT..... 615
    :|||:|||||.....:|||||
205 LeuGlySerGlySerGlnPheIleTyrLysGlySerArgTyrGlnLeu 221
    :|||:|||||.....:|||||
616 .....GATGATGACAAACACGCGGATTTATCTACTCGCGGCGAT 655
    :|||:|||||.....:|||||
221 uIleLeuThrGlnLysAspLysGlnLysLeuLeuArgAsn.....T 236
    :|||:|||||.....:|||||
656 GGTAAATGGCGGCAATACACATATCGAGGTTGGGGAATATATGCGTA 705

```

```
11 .....:
236 rpaervalglyasp...AsnleuglueuValglYasnAlatYrThr 251
706 NTAGTTGACGGC.....GATGGCCCATGCAACGAC..... 741
252 TyrllylealaglYThrProTYrLysValAsnHisgluAsnAsnlyLe 268
742 .....TATGGC..... 747
268 uileglypneHisserIySgluHisSerAspProLysglYlel 285
748 .....CCTATGCCATTGCGAGTGGCGGACGCGACGCGTTGC 786
285 euSerGlnAspProLeuThrAsnTYrAlaValleuglyAspSerGlySer 301
787 CCATGTTATTATGACAAACAAACAATTAATGGCTGCACAGCGGT 836
302 ProLeuPheValTYrAspArgluLysglYStrLeuPheLeugly.. 317
837 TTACAAACCGCTACCTTATTCGGCAGGAAACGGTTTCCAGCTGA 886
318 .....SerTYrAspPheTrpAlaGlyTYrAsn..... 326
887 TACGCAAGATGGTTTACGATGACATTTACAGAGCGGATACATACC 936
327 ..LysLysSerTrpGlnGluTrpAsnIleTYrLysHisgluPheAlaGlu 342
937 GTCATTTTGAACCGCGCAGTACGACATTTTCCTTTACATCCACAA 986
343 LysIleTYrGlnGlnTYrSerAlaGly.....SerLeuThrGlySerAs 357
987 C.....AACGGTACGGGTACGGTAAACAGAA 1012
357 nThrGlnTYrThrTrpGlnAlaThrGlySerThrIleThrGlyG 374
1013 CCACGAAAGGTNTCCATCCAAAGCTTAAAGTACACAGCTCCGACTG 1062
374 LysglYgluProLeu..... 378
1063 TTGACGATCTTTGATGAATGATATTAAGAACCACTTACGGCGCAG 1112
379 .....SerValAspLeuThrAspGly..... 385
1113 GGGGTATATCAGTACCGTCCAGGTTAAACACGCTGAACACTTCTT 1162
386 .....LysAspLysProAsnHisglYSerSerIleThrL 397
1163 TTATGATTACGCAACGCAACATCTTATCAACAACATCAACCA 1212
397 eu.....LysglYSerGlyThrLeuThrLeuAsnAsnHisIleAspGln 411
1213 GCGCGGGGGGTTGATTATTTGAAGTGAATTTTACGGTC...TCGCCGGA 1259
412 GlYAlaIaglyGlyLeuPhePhegluGlyAspTYrGluValLysglYThrSe 428
1260 AACCAACGAAACGTGGCAAGCGCGGGCTCATATCACTGACAGACAGTA 1309
428 rAspSerThrThrTrpLysglYAlaGlyValSerValAlaAspGlyLysT 445
1310 CGGTACTGTTGAAGTAACGGCGTGGCAACGCGCTGCCAAATC 1359
445 hValThrTrpLysValHisAsnProLysTYrAspArgLeuAlaIlyStle 461
1360 GGCAGAGCAGCTGCACGTTCAAGCAAGCAAGGGAACCAAGGCTCGAT 1409
462 GlYLysglYThrLeuValValGluGlyLysglYLysAsnGluGlyLeuLe 478
1410 CACGCTGGGCGACGTACAGCTATTTGGATCAGCAGGACAGCATAAAG 1459
478 uLysValGlyAspGlyThrValIleLeuLysGlnLysAlaAspAlaAsnA 495
1460 GCAGAAACACGCTTACTGAATCGGCTGTGTCAGGCGGCGGTACG 1509
495 snLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgSerThr 511
1510 GTGCAACTGATATGCCGATATATCATTTCAACCCGCAACACTCTATTGG 1559
512 LeuValIleAsnAspAspLysglYValAspProAsnSerIleTYrPheG 528
1560 CTTTCCGGCGGACGTTGGATTTAACGGGATTCGTTGCTGCCACC 1609
528 yPheArgGlyGlyArgLeuAspLeuAsnGlyLysSerLeuThrPheAspH 545
1610 GATTCGAAATATCCGATGAAGGCGGATGATTGNCATCATATATCCACA 1659
545 IStleArgsnIleAspAspGlyAlaArgValValAsnHisAsnMetThr 561
1660 ACACATCCACCGTTTACATTCACAGGAATGAATATATACCAACCG.. 1707
562 AsnThrSerAsnIleThrIleThrGlyGlySerLeuIleThrAspProAs 578
1707 ..... 1707
578 nThrIleThrSerTYrAsnIleGluAlaGlnAspAspAspHisProLeuA 595
1707 ..... 1707
595 rglIleArgSerIleProTYrArgGlnLeuTYrPheAsnGlnAspAsnArg 611
1707 ..... 1707
612 SerTYrTYrThrLeuLysLysglYAlaSerThrArgSerGluLeuProG 628
1708 ....AGGGTAAGAAAT..... 1719
628 nasnSerGlyGluSerAsnGlnAsnTrpLeuTYrMetGlyArgThrSera 645
1720 .....ATCAATAGACTTAATTACAGCAAGAAAT 1749
645 spGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsn 661
1750 GCCTACACCGGTTGTTGGCGAGAAAGATACGACCAACAAAG...AACGG 1796
662 GlyPheAsnGlyTYrPheGlyGluGluGluTYrLysAlaThrGlnAsnG 678
1797 GCGGCTCAACCTGTTTACCGCCCGCGGAGAGACCGCACCCGCTGCG 1846
678 LysLysLeuAsnValThrPheAsnGlyLysSerAspGlnAsnArgPheLeu 695
1847 TTTCCGGCGGACAAATTTAAACGCAACATCAACCAACAAACGCAAA 1896
695 euThrGlyGlyThrAsnLeuAsnGlyAspLeuAsnValGluLysglYThr 711
1897 CTGTTTTCAGCGGACGACGACGCGCAGCGCTACATCATTTAGGAAG 1946
712 LeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 728
1947 CGGCTGTCAAAATGAGAAAGT.....ATCCCAAGAGGAAATCGTGT 1990
728 eSerSerThrLysLysAspProHisPheThrGlnAlaAsnGluValVal 745
1991 GGCAGCAACGATGATCNACCGCACGTTTAAAGCGGAAATTTCCATAT 2040
745 alGluAspAspTrpIleAsnArgAsnPheLysAlaThrTrpMetAsnVal 761
2041 CAGGCGCGGCGGCGGATTC...CGCAATGTCCCAAGTGAAG 2087
762 ThrIlysnAlaSerLeuTYrSerGlyArgAsnValAlaAsnIleThrSe 778
2088 CGATTGNCATTTGACAAATCAGCCCAAGCAAGTTTGTGTGCGCACCGC 2137
778 rAsnIleThrLysSerAsnAsnAlaGlnValHisIleGlyTYr.....L 793
2138 ATCAAGCCATACATCTGTACACGTTTCGAGCTGACAGCTGCACAAAT 2187
793 yStrnGlyAspThrValLysValArgSerAspTYrThrGlyTYrValThr 809
```

[illegible]

```

1377 GlnAlaIysProGlnThrGluProAlaArgIuAsnValSerThrValAs 1393
3501 ..... 3501
1393 mThrLysGluProGlnSerGlnThrSerAlaThrValSerThrGluGlnP 1410
3501 ..... 3501
1410 roAlaLysGluThrSerSerAsnValGluGlnProAlaProGluAsnSer 1426
3501 ..... 3501
1427 IleAsnThrGlySerAlaThrThrMetThrGluThrAlaGluLysSerAs 1443
3501 ..... 3501
1443 PLYSPGInMetGluThrValThrGluAsnAspArgGlnProGluAlaA 1460
3501 ..... 3501
1460 snThrValAlaAspAsnSerValAlaAsnAsnSerGluSerSerGluSer 1476
3502 ..... CGTTATGCCAATAGCGGTTGAGTGAATTTCCGCCAC 3539
1477 LysSerArgArgArgSerValSerGlnProLysGluThrSerAlaG1 1493
3540 GGTCAACAGCGCTTTCCGCCGTACAGGAC..... 3567
1493 uGluThrThrValAlaSerThrGlnGluThrThrValAlaAspAsnSerValS 1510
3567 ..... 3567
1510 eThrProLysProArgSerArgThrArgSerValGlnThrAsn 1526
3568 ..... GAATTGACCGCGCTGTTCCGAGACGCGCGCA 3602
1527 SerTyrGluProValGlnLeuProThrGluAsnAlaGluAsnAlaGluAs 1543
3603 CGCN..... 3606
1543 nValGlnSerGlyAsnAsnValAlaAsnSerGlnProAlaLeuArgAsnL 1560
3606 ..... 3606
1560 eThrSerLysAsnThrAsnAlaValLeuSerAsnAlaValAlaLysAla 1576
3606 ..... 3606
1577 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisTleSerG1 1593
3607 ..... GTTGGACAGCAGCNGCATCC 3625
1593 nLeuGluMetAsnAsnGluGlyGlnThrLysAsnValThrPleSerAsnThrs 1610
3626 GGNACACCAACACTACCGTCCGCAAGATTCCGCCCTACCGCCACAA 3675
1610 eMetAsnLysAsnTyrSerSerGluGlnTyrArgArgPheSerSerLys 1626
3676 ACCGACCTGGCCAAATCGGTATGCAAAACCTCGGAGCGGG...CG 3722
1627 SerThrGlnThrGlnLeuGlyTyrPaspGlnThrIleSerAsnValG1 1643
3723 CGTGGCATCTGTTTCGCACACGCGACGCAAAACAMTTCGACGACG 3772
1643 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPhaAspLysA 1660
3773 GCATCGGCAACTCGCACGGCTTGCCACGGCGCCCTTTTCGGGCAATAC 3822
1660 LaserSerLysAsnThrLeuAlaGlnValAlaAsnPhaTyrSerLysTyr 1676
3823 GGCATCGGCGAGTTCCGACATCGGCATCAGCGGCGGCTTTAGCAG 3872

```

```

1677 AlaAspAsnHisThrPtyrLysGlyIleAspLeuGlyTyrGlyLysPheG1 1693
3873 CGGCANTCTNTTCAGACGCGATCGGACGCAAAATCCGCCCGCGCTGCTCC 3922
1693 nSerAsnLeuGlnThrAsnAsnAsnAlaLysPheAlaArgHisThrAlaG 1710
3923 ATTACGGCATTCAGGACGACGATACCGCGCGCTTCCGGCGGATTCGGCAGTC 3972
1710 InIleGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheAlaVal 1726
3973 GAACCGTACATCGGCGCAACGCGCTATTTGTCGCAAAAGCGGATTACCG 4022
1727 LysProThrValGlyValAlaArgTyrSerTyrLysSerAsnAlaAspPheAl 1743
4023 CTACGAAACGTCATATCGCCACCCCGCTGCTTCGCTTCACCGTACCG 4072
1743 aLeuAlaGlnAspArgIleLysValAsnProIleSerValLysThrAlaP 1760
4073 GNGCGGCATTAAAGCGCATTTATTCATTCAAACCGCGCACACATNTCC 4122
1760 heAlaGlnValAlaPLeuSerTyrThrHisLeuGlyGlu...PheSer 1775
4123 ATCAGCCCTTATTATTTNAGCCTGTCTATACCGATGCGGCTTCGGCAAGT 4172
1776 IleThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnG1 1791
4173 CCGAACACGCGTCATATCCGCGNGTATGCTCAGAGATTTGGCAAAACCC 4222
1791 LysIleAsnValSerValTyrAspPheAlaTyrAspValAlaGlnAsnG1 1808
4223 GCACTGGCGAATGGCGTAAACCGCGCAATCAAGGTTTCACGCTGTC 4272
1808 InGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1824
4273 NTCCAGCGTCCGCCGCCAAAGNCCGCAACTGACGACGACACACACCGC 4322
1825 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1841
4323 GGCATCAATTAAGGCTAC 4341
1841 aGluValLysLeuSerPhe 1847

seq_name: p1r2:A81937

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) MNA0905 [imported] - Neisseria meningitidis
N:Alternate names: Iga proteinase; Iga1 proteinase (EC 3.4.21.7) [misnomer]; Immunog1
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain HF159; strain SM1027
C:date: 05-May-2000 #sequence: revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parikh, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jørgensen, K.; Leather, S.; Mould, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; M0ID:20222556
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737
A:Experimental source: serogroup A, strain 22491
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; M0ID:95302961
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 33-548 <LOW>
A:Cross-references: EMBL:X82470; NID:g732854; PIDN:CAA57853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown

```


A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82471; NID:q732858; PIDN:CAAS7854.1; PID:q732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82472; NID:q732852; PIDN:CAAS7855.1; PID:q732853
A:Experimental source: strain SM1027
C:Genetics:
C:Gene: Iga: NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 1410.00 Length: 1849
Ratio: 1.495 Gaps: 60
Percent Similarity: 51.001 Percent Identity: 26.501

alignment_block:
US-09-303-518D-651 x A81937

Align seg 1/1 to: A81937 from: 1 to: 1773

```
49 AAACCGCCGATCCGCTTCTGCGCTGCTTACTTACGATATGCGCTGTC
||||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATTTCTCCCAAGCTTGGCGGAGACACTTATTTGGCCATCA 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18 atyrAlaLeuThrProTyrSerGluAlaLeuValArgAspAspValA 35
149 ACTACCAATACTATGCGCAGTTCGCCGAAATTAAGCAAGTTCAGCTC 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
35 sptyGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheVal 51
199 GGGCGGAAAGATATTGAGTNTACAAACAAAGGGAATGGTGGCGCA 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 GlYAlaThrAspLeuSerValLysAsnLysGlnIleAsnIleGlyAs 68
249 ATCATGTACAAAGCCGCGATGATGATTTTCTGTGCGCTGAC. 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85
298 .GGCGTGGCGCATTTGGTGGCGCATATATATTTGTGCGCGCATC 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 rGArgThrLeuThrValIleAspProGlnIlyrAlaValSerValLysHis 101
346 AAGCGCGCTATACACAGCTTGATTTGGTGGGAGAGAAAGAAATCCGCA 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 ValLysGlyAspGluIleSerTyrTyrGlyHisAsnGlnIleLys 118
396 T.....CAGCACCGTTTCTTACCAATTTGAAAAAATAATATA 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 pValSerAsnAspGluAsnGlnIlyrArgSerValAlaGlnAsnAspTyrG 135
440 AGCCTGACAAATCA.....CACCTTACAAAGCG.....GATTAN 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 IuproAsnLysAsnThrHisHisGlnLysGlnIlyrArgLeuGluAspTyr 151
475 CATATGCGCGCTTGATTAATTTGTGACAGATACACAGCTGCGCAAT 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaP 168
525 GACGAGTGACATGAGGGG...AATACCTATTCGATTAAGAAATAATC 571
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 oThSerAlaGlyLysGlyValGluThrTyrLysAspLysAsnArgPheS 185
572 CCGAGCGCTGCCGATCGCTCAGACACACTATTTGGCTTATGATGAT 621
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 erGluPheValArgValGlyAlaGlyThrGlnPheGlnIlyrAsnSerArg 201
```

```
622 GACAAACAGCGGATTTCTCTACTCCGCGCATGTTATTTGGCGGCA 671
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 TyrAsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyTh 218
672 TACACATATGCGAG.....GATTANTTAGTTGACGCGGATGTCGCCAT 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 rProTyrGlnAspValAsnValThrSerAsnLeuAsnGlnIlyrLeuI 235
685 .GGTTGGGGAATTAATGCGCTANTTACTTTGACGCGGATGTCGCCAT 732
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 IeGlyPheGlyAspAsnSerLysHisHisSerProGlnIlyrLeuLysGlu 251
733 GCGAACGACTATGGCCCTATGCCATTCGAGTGGCGGAGCGGACGCG 782
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ValLeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAspSerG 268
783 TTCGCGCATGTTTATTTATGACAAACAAACATTAATGCGCTGCTCAC 832
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
268 ySerProLeuPheAlaTyrAspLysGlnIlyrArgTyrAlaPheLeuG 285
833 GAGTTTACAA.....ACGCGTACCTTATTCGCGGAGGAAACGCT 876
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 LyAlaTyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSer 297
877 TTCGAG.....CTGATACGCAAGATTTGTTACATGACATTTACAG 920
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 TrpGlnIlyrAsnIleTyrLysGlyLysGluPheLysIlyrLeuLysG 314
921 AGCGATACACATACCGCTTNTTGAACCGCGGAGTAAAGGACATTTT 970
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 nArgAsp...AsnAlaGlyThrIleLysGlyTyrGlyGlnHisHisTrpL 330
971 CTTTATATCCACACACACGCTACGCGTACGCTACGCAACCAAGAA 1020
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 yThrThrGlyThrAsnSerHisIleGlySerThr..... 341
1021 AAGSTNTCCATCCAAAGCTTAAGTACACAGATCCGATGTTTGACGA 1070
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 .....AlaValArgLeu..... 345
1071 ATCTTGAATGAACCTGATTAAGAACACGATTACGCGGAGGGGTGTTA 1120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 .....AlaGlyAsn.... 348
1121 ATCAGTACCGTCCAAGTTAAACACGCTGAACCTTTCTTTATCGAT 1170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe..... 360
1171 TACGCAACGCAACATCATCTATCAACACATCAACCAAGCGCGGG 1220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnIlyrAla 377
1221 CGCTTTGATTTTGAAGTGATTTTACGGTTCGCGCTGAAACCAAGAA 1269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
377 yGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyI 394
1270 .ACGTGGCAAGGCGCGGCTTCATATCAAGTGAAGACGATACCTTACT 1317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
394 IeThrTrpLeuGlyAlaGlyIleAspValAlaAspLysLysValVal 410
1318 TGGAAAGTAAACGCGCTGCGCAACGACCGCTGCTCCAAATGCGCAAG 1367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 TrpGlnValLysAsnProAsnGlnLysAspArgLeuAlaLysIleGlyLys 427
1368 CACGCTGCACGTTCAAGCCAAAGGGGAAACCAAGGCTGATCAACGTTG 1417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
427 yThrLeuGlnIleAsnGlnIlyrThrGlyValAsnGlnIlyrLeuLysValG 444
1418 GCGACGCTACAGTCAATTTGATCAGACGAGGAGAGATTAAGGCAAAA 1467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
444 LyAspLysThrValIleLeuAsnGlnIlyrAlaAspAlaAspLysLysVal 460
```



```

1053 hrArgLInLysAlaGluInGluInLysSerSerAlaGluLeuAlaArgArg 1069
2748 GCAAAACGGGAGNGTGTACAGACACCGCGCGCGCGCGCGCGTCC 2797
1070 HisGluLysGlu.....ArgGluAlaAlaGluLeuSerAlaLysGlnLys 1084
2798 TATT.....ATCGGTTACACGGCCA 2817
1084 sValGluAlaGluArgGluAlaGlnAlaLeuAlaValaArgArgLysAlaG 1101
2818 ACTTCGGTAGACATCCGTTTCACACCGCTGACGGTAAACGGCAATTGAA 2867
1101 LuAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaArgArgHisGlu 1117
2868 CNGTCAAGAACATTCGCTTTATGTGGAACTTTCGGCTA..... 2909
1118 LysGluArgGluAlaAlaGluLeuSerAlaLysGlnArgValaGluGlu 1134
2910 ..CCGAACGCAAAATTGAACTGGCGGAAAGTTCCGAAGGNACTTACAC 2958
1134 uGluArgArgGlnThrAlaGlnSerGlnProGlnArgArgLysArgArg 1151
2959 TTGGCGGTCAACAATAC.....CGGCAAGCAACCGGT 2990
1151 LaAlaProGlnAspTyrMetAlaAlaSerGlnAspArgProLysArgArg 1167
2991 AAGCTTCATCAATTTGACGGTAGTGAAGGAA..... 3023
1168 GlyHisArgSerValGlnGlnAsnAsnValGluLeuAlaGlnAlaGln 1184
3024 .....AGACACAA..... 3032
1184 aGluLeuAlaArgArgGlnGlnGlnGluArgLysAlaAlaGluLeuLeu 1201
3033 .....ACCGCTGTC 3041
1201 LaLysGlnArgAlaGluAlaGluArgGlnAlaGlnAlaLeuAlaAlaArg 1217
3042 CGAAACCTTAATTTCACCCCTGCAGAAACGACAGCTGATGCCGCGCGCT 3091
1218 ArgLysAlaGluAlaGlnAlaLysArg..GlnAlaAlaGluLeuAlaGln 1234
3092 GGGTTACCACTCATCCGCAAGAGCGGAGCTTCGCGCTGCATTAAT... 3138
1234 IsArgGlnGlnGluAlaGluArgLysAlaAlaGluLeuSerAlaAsnGlnLys 1250
3139 ...CCGCTCAAGAACACAGAGCTTCCGACAACTCGCAGGACGAGAAC 3185
1251 AlaAlaAlaGluAlaGlnAlaLeuAlaAlaArgGlnGlnLysAlaLeuAl 1267
3186 CAAAAACAGCGCGGAAAAAGACACCGCGCAAGCCTT..... 3222
1267 aArgGlnGlnGlnGluAlaArgLysAlaAlaGluLeuAlaValLysGlnL 1284
3223 .....GACGGCTGATTCGGCGCGCGCGCGCATGCC 3252
1284 yAlaGluThrGluArgLysThrAlaGluLeuAlaLysGlnArgAlaAla 1300
3253 GCCGAA.....AAGACGAAAGCGTTGCCGAGACCGC 3284
1301 AlaGluAlaAlaLysArgGlnGlnGlnAlaArgGlnThrAlaGluLeuAl 1317
3285 CCGGNGCGGAGCGGGAATAATGTGGCATTTGACGCGGAGAA..... 3330
1317 aArgArgGlnGluAlaGluArg.....GlnAlaAlaGluLeuLys 1330
3331 ..GAGAAAAACGGGTGCAGCGGATTAAGACAGCGCAGCTTGCAGAAACAG 3378
1330 eValaLysGlnLysAlaGlnThrAspArgGluAlaAlaGluSerAlaLys 1346
3379 CGCGAAGCGGGAACCGCGCGGNTACACCGCCTTC..... 3414
1347 ArgLysAlaGluGlnGluGlnHisArgGlnAlaAlaGlnSerGlnProGln 1363
3414 ..... 3414
1363 naArgArgLysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGln 1380
3415 .....CCCGCGCGCGCGCGCGCGCGCGGATTTGCCGACCG... 3453
1380 snArgProLysArgArgGluArgArgSerThrLeuProAlaProProSer 1396
3453 ..... 3453
1397 ProSerPheAspSerSerAlaTyrAlaAlaProAlaAlaLeuHisAsnPr 1413
3453 ..... 3453
1413 asPrrTyrGluAsnAspTyrGluGluLeuProLeuAspAlaLeuGlu 1430
3453 ..... 3453
1430 sPGLuAsnValSerGluSerValaAspThrSerAspLysGlnProGlnAsp 1446
3453 ..... 3453
1447 AsnThrGluLeuHisGluLysValaGluThrValaSerLeuGlnProAlaGln 1463
3454 ...CAGCC...CAACCGCAACCTCAACCCCAAC 3481
1463 aAlaGlnProAlaGluAlaGlnAlaAlaThrGlnProGlnAlaGlnAlaAla 1480
3482 CGCAGCGGACCTGATNAGCGGTATGCCATACCGGTTTGAAGAAATT 3511
1480 laGlnAlaAspAlaValaSerThrAsnThrAsnSerAlaLeuSerAspAla 1496
3532 TCAGCCAGCTCAACAGCTTTTCGCC...GTACAGAGCAATTTGAGACG 3578
1497 MetAlaSerThrGlnSerThrLeuLeuAspThrGlnAlaAlaSerLeuThr 1513
3579 CGGTTCCTCGAAGACCGCCG...AACGCGTTTGA 3613
1513 gHisThrLeuGlnLysSerArgAlaAspAlaGluLysAsnSerValaLys 1530
3614 CAAGCGCATCCGCGNACACAAACATACCTGCTGCAAGATTCGCGGCC 3663
1530 etSerAsnThrGluGluArgAspTyrAlaSerAlaGlnTyrArgArg 1546
3664 TACCGCCACAAACCGACCTGCGCAATCGGTGAGAAAACTCGG 3713
1547 PheSerSerLysArgThrGlnThrGlnThrGlnThrLysLeuAspArgSerLeu 1563
3714 CAGCGGG...CGGTCGGCATCCGTTTTCGCAACACCGGACGCAAAACA 3760
1563 rGluAsnMetGlnThrGluGluValaLeuThrTyrSerAspSerThrHis 1580
3761 NCTTCAGACGAGCATCGCAACTCGGACAGCGCTTCCGCGCGCGCTT 3810
1580 hrPheAspGlnAlaSerGlnLysAsnThr...PheValGlnAlaAsnLeu 1595
3811 TTGCGGAATACGGCATCGGACAGTTGCACATCGGCAATCGAGCGGCGC 3860
1596 TyrGluLysTyrTyrLeu.....AsnAspAl 1604
3861 GGGTTTACGAGCGCGCANTCTNTCAGACGCGCATCGAGGCAAAATCGCG 3910
1604 aTrrTyrMetAlaGluAspThrLeu.....AlaGlySerLeuArg 1618
3911 CGCGCGTGTGCATTTACGGCATTCAGGACAGATACCGCGCGGTTTC... 3957
1618 eArg.....LeuGlnThrGlnGlnLysAlaAsnPheAsn 1629
3958 .....CG 3959
1630 ArgThrSerThrLeuThrGluLeuThrLeuGluAsnThrLeuLysLeu 1646

```

```

3960 CGGATTCGCATCGACCGCTACATCGCGCAACGCGCTATTGTCGCAA 4009
      : ||| ||| ||| ||| : : : : :
1646 nglpnlhgluilevalproseralaglileargtyrserarleusers 1663
      : : : : :
4010 AAGGGGATTCAGCTACGAAACGTCATATMCGCACCCCGGCTTCGCG 4059
      : : : : :
1663 erlialaspyrlylsleuglyasnaspservallysvalserSerMeSer 1679
      : : : : :
4060 TTCAACCGNTAACGCGCGGCGATTAAAGCGAATATTCATTCACACCGC 4109
      : : : : :
1680 VallysrThrleuthrThrIaGlyLeuasphealatyrrarphelys...Va 1695
      : : : : :
4110 GCACACATNTCCATCACNCTTTTNNAGCGTCTCTAT...ACGGATG 4156
      : : : : :
1695 lGlyasnleuthrvallylsproleuSerAlaIaIaIaIaIaIaIaIaIa 1712
      : : : : :
4157 CCGCTTCGGCGAAAGTCGCAACAGCGCTCATATACCGGNTATTCGCTCAG 4206
      : : : : :
1712 YrGlylysglyGlyalasnvalGlyGlyasnserphevalTyrlysaIa 1728
      : : : : :
4207 GATTTCCGCAAAACCCGCGATGCGGAATGGGCGTAAACGCCGAATCAA 4256
      : : : : :
1729 AspaenglnglnglntyrSerAla.....GlyAlaIaIaIaIaIaIaIaIa 1743
      : : : : :
4257 AGCTTTCACGCTGTCNCTCCACGCTCGCGCGCAAGGCGGACGCTCGG 4306
      : : : : :
1743 gasnvalThrleuasnvalasnGlySerTleThrlysglylysglnleug 1760
      : : : : :
4307 AAGCGCAACAGACGCGCGGCGCATCAATTAAGCTACCCGCTGG 4347
      : : : : :
1760 lulysglnlysserGlyglnleuIleugIleargPhe 1773

```

seq_name: p1r2:C81169

seq_documentation_block:

```

IGA-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [Imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: strain MC58 serogroup B; strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Retelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.R.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignanl, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; M01D:20175735
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <DET>
A:Cross-references: GB:AE002424; GB:AE002098; NID:97225923; PIDN:AAF41117.1; PID:9722592
A:Experimental source: serogroup B, strain MC58
R:Lomholt, H.; Poulsen, K.; Mogensen, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseri
A:Reference number: S61314; M01D:95302961
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82477; NID:9732856; PIDN:CAAS7860.1; PID:9732857
C:Genetics:
A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase, metalloproteinase

```

```

alignment_scores:
  Quality: 1404.00      Length: 1878
  Ratio: 1.479          Gaps: 60
  Percent Similarity: 50.532  Percent Identity: 26.198
alignment_block:

```

```

US-09-303-518d-651 x C81169
Align seg 1/1 to: C81169 from: 1 to: 1815

```

```

49 AAACCGCGCGCATCCGCTTCGCTGCTTACTTACCATATGCGCTGC 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrlysrGlypheIysIleasnAlaIleSerleuSerIlepheIeVal 18
      : : : : :
99 GTTCGGCATTCCTCCCAAGCTTGGCGGACACACATTCATTTCGGCATCA 148
      : : : : :
18 arYrAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 35
      : : : : :
149 ACTACCAATATATCGCGACTTGTGCGGAAATTAAGCGAATGTTCCAGTC 198
      : : : : :
35 sptyrGlnIlepheIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 51
      : : : : :
199 GGGGCGAAAGATATTCAGTNTACAAACAAAGGAGGAGTTCGCGGCA 248
      : : : : :
52 GlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 68
      : : : : :
249 ATCAATGACAAAGCCCGATGATGATTTCTGTGTGTCGCGTAAAC. 297
      : : : : :
68 nAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 85
      : : : : :
298 ..GGCGTGGCGGATTCGTGGCGATCAATATATTCGATGCGCAT 345
      : : : : :
85 rgarGthrleuthrValIleasproGlnTyrAlaIaIaIaIaIaIaIaIa 101
      : : : : :
346 AACGGCGGCTATACAAACGTGATTTGGTGGCGGAAGAAATATCCGA 395
      : : : : :
102 VallysglyaspIuIleSerTyrTyrGlyIaIaIaIaIaIaIaIaIaIa 118
      : : : : :
396 T.....CAGCACCGTTCCTTACCAATTCGAAAGAAATATATTA 439
      : : : : :
118 pValserAsnaspIuasnIuTyrArgSerValaIaIaIaIaIaIaIaIa 135
      : : : : :
440 AGCGCTGCAATTC.....CACCTTACACGCG.....GATTAN 474
      : : : : :
135 luproasnlysaSnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 151
      : : : : :
475 CATATGCGCGGCTTCGCTAATTCGACAGATGCAACGCTGCGCAAT 524
      : : : : :
152 AsnmetAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 168
      : : : : :
525 GACGAGTACATGAGGGG...AATACCTATTCGATTAAGAAATATATC 571
      : : : : :
168 oThrSerAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 185
      : : : : :
572 CCGAGCGTTCGCGATCGGCTCGAGCACACCATATTCGCTATGATGAT 621
      : : : : :
185 erGluIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 201
      : : : : :
622 GACAAACAGCGGATTCATCTACTCCGCGCATATGTTATTCGCGGCAA 671
      : : : : :
202 TyrIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 218
      : : : : :
672 TACACATATGACG..... 684
      : : : : :
218 rProTyrGlnaspValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 235
      : : : : :
685 ..GGTGGGAAATATATGCGGTANTTATTCGACGCGGATTCGCGCAT 732
      : : : : :
235 lGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 251
      : : : : :
733 GCCAAGCATATGCGCCTATGCGATTCGAGGTGCGGACGCGACGCGG 782
      : : : : :
252 ValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 268
      : : : : :
783 TTGCGCAATGTTTATATATGACAAACAAATTAATGATGCGTGCATCAG 832
      : : : : :
268 ySerProIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 285
      : : : : :
833 GAGTTTACAA.....ACCGCTACCTTATTCGCGGACGAGAAACGCT 876

```



```

2386 .....GTGGCAATGCCCAAGCACATTTATACACCCATTAACGG 2429
      |||
854 AsnAsnIaSerAlaGlnSerAlaAsnIyStrIyThrIleuLysII 870
      |||
2430 CAACNCATCGGNTTCGGGCATGCTTCATTTATCTAAGCAACACGCCG 2479
      |||
870 eAsn...HisIeuSerGIyAsnGIyHisPheHisTyrIleuThrHisIeuA 886
      |||
2480 CACAAAC..... 2487
      |||
886 IalYAsnIleuGIyAspIyValIeuValIyGlnSerAlaSerGIyHis 902
      |||
2488 .....GGCAGTCT 2495
      |||
903 TyrGIleuHisValGlnAspIyStrGIyIuProAsnGIyGluGIyLe 919
      |||
2496 GAGCGTTCCGACAAACGCT.....AAGCAAAACGTAAGCCATT 2533
      |||
919 uAsnIeuPheAspIaSerSerValGlnAspArgSerArgIeuSerValS 936
      |||
2534 CCGCAGCTAACGGCAATGTCCTCCATAGCCGATTAAGCAGATTCGATT.. 2581
      |||
936 erIeuAlaAsnHisValAspIeuGIy..AlaIeuArgTyrThrIleTy 952
      |||
2581 ..... 2581
      |||
952 sThrGIuAsnGIyIleThrArgIeuTyrAsnProTyrAlaGIuAsnArg 969
      |||
2582 .....TTGAAAACAGCCGCTTTACCGCAACT..... 2609
      |||
969 rGArgValIySerAlaProSerPro..AlaThrAsnThrAlaSerGlnAl 985
      |||
2609 ..... 2609
      |||
985 aGlnIyAlaThrGlnThrAspGIyAlaGlnIleAlaIySProGlnAsnI 1002
      |||
2610 .....CAG 2612
      |||
1002 IeValValAlaIaProProSerProGlnAlaAsnGlnAlaGIuGIuAlaIyS 1018
      |||
2613 CGGCAGCAAGAMACAGCATTAACCTTAAGACAGCGAATGGAC..... 2657
      |||
1019 ArgGIuGlnAlaIySAlaGIuGIuValIySArgGIuGlnAlaGIuAlaGI 1035
      |||
2658 .....GCTCCGTCAGCGACGGAA 2676
      |||
1035 uArgIySerAlaGIuIleuAlaIySGIuSAlaGIuAlaGIuArgIuA 1052
      |||
2677 TTAGCAATTTAAACCTTGACACGCCACCATTAACATTCGCGCTA 2726
      |||
1052 IaArgGIuIeuuAla.....ThrArgGIuIySAlaGIuGIuArgSer 1066
      |||
2727 TCGCCACGATGCTGCAGCGCGCAACCGCGAGNGTGCAGACAGCGCG 2776
      |||
1067 SerAlaGIuIleuAlaIaArgIArgHisGIuSGIu.....ArgIuAlaAl 1081
      |||
2777 GCGCGCGTCCGCGCTCCCTATT..... 2801
      |||
1081 aGIuIeuSerAlaIySGIuSAlaGIuAlaGIuArgIuGIuAlaGIuAla 1098
      |||
2802 .....ATCCGTTACACCGCAACTTCGTAGAAATCCGTTTCAACAGCT 2846
      |||
1098 euaIaValaIaArgArgIySAlaGIuAlaGIuAlaIySArgGIuAlaAla 1114
      |||
2847 GAGCGTAACCGCAATTTGACNGTCAAGAACATTCGCTTATGTGCG 2896
      |||
1115 GIuIleuAlaIaArgHisGIuIySGIuSAlaGIuAlaGIuIeuSerAl 1131
      |||
2897 AACCTTTCGCGCTA.....CCGAAGCGACAAATTTGAAGCTGCGCGAA 2937
      |||
1131 aIySGIuIaArgValGIyGIuGIuGIuIaArgArgGlnThrAlaGIuSerGlnP 1148

```

```

2938 AGTTCGAAGNACTTACACCTTGGCGGTCAACAATAC..... 2975
      |||
1148 roGlnArgArgIySAlaArgAlaIaProGlnAspTyrMetAlaIaIaSer 1164
      |||
2976 .....CGCAACGAACCGGTAAGCTCGATTCATTTAGAGGTAGCGAAG 3019
      |||
1165 GlnAspArgProIySArgIArgGIyHisArgSerValGIuGlnAsnAsnVa 1181
      |||
3020 GGAA.....AGACACAA..... 3032
      |||
1181 IGIuIleAlaGIuAlaGIuAlaGIuIleuAlaIaArgArgGIuGIuGIuAla 1198
      |||
3032 ..... 3032
      |||
1198 rGIySAlaIaGIuIleuAlaIySGIuIaArgAlaGIuAlaGIuArgIu 1214
      |||
3033 .....ACCGCTCCGGAACCTTAATTTCACCCGCAAAACG 3070
      |||
1215 AlaGIuAlaIeuAlaIaIaArgArgIySAlaGIuAlaGIuAlaIySArg 1231
      |||
3071 AACACGTCGATGCCGCGGTGCGCTTACCACTCATCCGCAAAAGACGCG 3120
      |||
1231 g..GIuAlaIaGIuIleuAlaIaHisArgGIuGIuAlaGIuArgIySAlaIa 1247
      |||
3121 GAGTTCGCGCTGCATAT.....CCGTCGAAGAACAGACCTTCCGA 3164
      |||
1248 GIuIeuSerAlaIaSnGIuIySAlaIaIaGIuAlaGIuAlaIeuaIaI 1264
      |||
3165 CAACATCGGACAGCGAGGCAAGCAAGCGGCAAGCAAGCAAGCGCGC 3214
      |||
1264 aArgGIuGIuIySAlaIeuaIaIaArgGIuGIuGIuAlaIaIySAlaIa 1281
      |||
3215 AAAGCCTT.....GACGCGCTG 3231
      |||
1281 IaGIuIeuAlaValIySGIuIySAlaGIuThrGIuArgIyStrAlaGIu 1297
      |||
3232 ATTGCGCGCGCGCGCATGCCGCCGAA.....AAGAC 3263
      |||
1298 IeuAlaIySGIuIaArgAlaIaIaGIuAlaIaIySArgGIuGIuAla 1314
      |||
3264 AGAAGCGTTCGCAACCGCGCCGCGAGCGGAGGGAATAATGTGCGCA 3313
      |||
1314 aArgGIuThrAlaGIuIleuAlaIaArgArgGIuGIuAlaGIuArg..... 1328
      |||
3314 TTATGACGCGGAGGAA.....GAGAAAAAACGGGTGCAGCGGATTA 3357
      |||
1329 .....GlnAlaIaGIuIeuSerAlaIySGIuIySAlaGIuThrAspArg 1343
      |||
3358 GACAGCGCCTTGGGAACAGCGGCAAGCGGAAACCGCGGNTACAC 3407
      |||
1344 GIuAlaIaGIuSerAlaIySArgIySAlaGIuGIuGIuIuHisArgGI 1360
      |||
3408 CGCCTTC..... 3414
      |||
1360 naIaAlaGIuSerGIuProGlnArgArgIySArgArgAlaIaIaProGlnA 1377
      |||
3415 .....CCCGCGCGCGCGCGCGCGCGCG 3438
      |||
1377 sPtyrMetAlaIaIaSerGlnAsnArgProIySArgArgGIyArgArgSer 1393
      |||
3439 GATTTGCCGCAACG..... 3453
      |||
1394 ThrIeuProAlaProProSerProSerPheAspSerAlaTyrAlaAl 1410
      |||
3453 ..... 3453
      |||
1410 aProArgAlaIeuHisAsnProAspTyrGIuAsnAspTyrGIuGIuI 1427
      |||
3453 ..... 3453
      |||
1427 IeProIeuAspAlaIeuGIuAspGIuAsnValSerGIuSerValAspThr 1443
      |||
3454 .....CAGCCCAA..... 3462

```

```

1444 SerAspLysGlnProGlnAspAsnThrGlnLeuHisGlnLysTyrGlnAs 1460
1462 .....
1460 masPyrGlnGlnIleProLeuAspAlaLeuGlnAspGlnAspValSerg 1477
1474 .....
1474 LuserValAspThrSerAspLysGlnProGlnAspAsnThrGlnLeuHis 1493
1493 .....
1493 GlnLysValGlnThrValSerLeuGlnProArgAlaAlaGlnProArgAl 1510
1510 .....
1510 aGlnAlaAlaThrGlnLeuGlnAlaGlnAlaAlaAlaGlnAlaAspAlaVal 1527
1527 .....
1527 TNAAGCGTTATGCAATAGCGGTTTGAATTTCCGCGACCGCTCAAC 3546
1543 .....
1543 alSerThrAsnThrAsnSerAlaLeuSerAlaMetAlaSerThrGln 1543
1543 .....
1543 ACCGTTTCCGCC...GTACAGAGCAATTTGACCGCGTGTTCGCGAAGA 3593
1560 .....
1560 SerIleLeuLeuAspThrGlnValAlaSerLeuThrArgHisIleAlaGlnLys 1560
1560 .....
1560 CGCGCCG...AACGNGTTTGACAGCAGCGCATCCGCGN 3628
1560 .....
1560 sSerThrAlaAspAlaGlnLysAsnSerValAlaThrMetSerAsnThrGlyT 1577
1577 .....
1577 ACACCAACACTACCGCTTCCGAGATTTCCGCGCTACCGCCACCAAC 3678
1593 .....
1593 YrGlyArgAspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysArg 1593
1593 .....
1593 GACCTGGCCCAATGCGGTATGAGAAAACCTCGCGACGCGG...CGCGT 3725
1610 .....
1610 ThrGlnThrGlnIleGlyIleAspArgSerLeuSerGlnAsnMetGlnI 1610
1610 .....
1610 CGGCACTCCTGTTTTCGACAAACCGGACGCAAAACACTTTCGACGAGCA 3775
1627 .....
1627 eGlyGlyValLeuThrThrYrSerAspSerGlnHisThrPheAspGlnAlaAs 1627
1627 .....
1627 TCGGCACTCGGCGACGCGCTTCCGCGCGCGCTTTCGCGCAATACGCG 3825
1642 .....
1642 eGlyLysAsnThr...PheValGlnAlaAsnLeuTyrGlyLysTyr 1642
1642 .....
1642 ATCGGC...AGTTTCGACATCGGCGATCAGCAGCGCGCGGTTTTCAGCAG 3872
1659 .....
1659 LeuAsnAspAlaIlePyrValAlaIleAspIleGlyAlaGlySerLeuAr 1659
1659 .....
1659 CGGCACTTCATTCAGCAGCGCATCGGACAAATCCGCGCGCGCGCGC 3922
1676 .....
1676 gSerThrGlnGlnThrGlnGlnLysAlaAsnPheAsnArgAlaSerIleG 1676
1676 .....
1676 ATTACGCGATTCAGGCGACGATACCGCGCGTTCGCGCGATTCGCGATC 3972
1692 .....
1692 InThrGlnLeuThrLeuGlnLysAsnThrLeuLysIleAsnGlnPheGlnIle 1692
1692 .....
1693 ValProSerAlaGlyIleArgTyrSerArgLeuSerAlaAspTyrLys 1709
1709 .....
1709 CTACGAAACGTCATATCGCCACCGCGCTTCGTTCAACGCGNATAC 4072
1709 .....
1709 sLeuGlnLysAsnAspSerValLysValSerSerMetSerValLysThrLeuT 1776
1776 .....
1776 GNGCGGCGATTAAAGCAGATTATTCATTCAAACGCGCGCAACAGATTTC 4122
1791 .....
1791 hrAlaGlyLeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuThr 1741
1741 .....
1741 ATCAGCTTATTTAGCTGTCTAT...ACCATGCGCGCTTGGGCGAA 4169
4169 .....

```

```

1742 ValLysProLeuLeuSerAlaAlaTyrPheAlaAsnTyrGlyLysGly 1758
4170 AGTCGGAACACGCGTCAATACCGCGNATTTGGCTCAGAGATTTCGCAAAA 4219
1758 yValAsnValGlyLysAsnSerPheValTyrLysAlaAspAsnGlnG 1775
4220 CCCGCACTGCGGAAATGGGCGTAAAGCGCGAATTCAAAGGTTTCAGCGT 4269
1775 InTyrSerAla.....GlyAlaAlaLeuLeuTyrArgAsnValThrLeu 1789
4270 TCCNTCCAGCGTCGCGCGCGCAAGCGCAACCTGAGCGACGCAACAG 4319
1790 AsnValAsnGlySerIleThrLysGlnLeuGlnLysGlnLysSe 1806
4320 CGCGGCGATCAATATAGCTACCGCTCG 4347
1806 rGlyGlnIleLysIleGlnIleArgPhe 1815

```

seq_name: p1r2:154632

```

seq_documentation_block:
  tsh protein - Escherichia coli
  C:Species: Escherichia coli
  C:date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Dec-2000
  C:Accession: 154632
  R:Providence, D.L.: Curtiss, R.
  Infect. Immun. 62, 1369-1380, 1994
  A:title: Isolation and characterization of a gene involved in hemagglutination by an
  A:reference number: 154632; M01D:94178945
  A:Accession: 154632
  A:status: preliminary; translated from GB/EMBL/DBJ
  A:molecule type: DNA
  A:residues: 1-1377 <RFS>
  A:cross-references: GB:127423; NID:9469235; PIDN:AAA24698.1; PID:9469236
  C:superfamily: Iga-specific metalloendopeptidase

```

```

alignment_scores:
  Quality: 1084.50      Length: 1508
  Ratio: 1.378          Gaps: 54
  Percent Similarity: 52.188      Percent Identity: 25.995

```

alignment_block:
US-09-303-518D-651 x 154632 ..

Align seg 1/1 to: 154632 from: 1 to: 1377

```

90 ATGCGTGTGCTTCGCGATTCCTCCCAACCTTGGCGGACACACTTATT 139
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  :
42 IleProValLeu...PheSerAlaGlySerLeuAlaGlyThrValAsnA 57
140 TCGGCATCAACTACCAATACTATCGCGACTTTCGCCAAATAAAGCAAG 189
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  :
57 snGluLeuGlyTyrGlnLeuPheArgAspPheAlaGlnLysGlyMet 73
190 TTTCAGTTCGGGCGAAAGATATTGAGGTTCATACAAACAAAGAGGAGT 239
   ||| ||||| :|||||:|||||:|||||:|||||:|||||:|||||:  :
74 PheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGlnGlyIle 90
240 GGTGCGCAATTCATGACAAAGCCCGGATGATTGTTTCTGTGTG 288
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  :
90 eValGly...ThrLeuAspLysAlaAlaMetProAspPheSerAlaValA 106
289 ..TCGCGTAAAGCGCGTGGCGCATTTGGTGGCGCATATATATTGAC 336
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  :
106 sPserGlnIleGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSer 122
337 GTGCGCATATAGCGCGGTATATACAGGTGATTGTTGGCGGAAGAG 386
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  :
123 ValLysHisAsnGlyLysTyrThrAsnValSerPheGly...AspGly 138
387 NAATCCGATCAGCACACGTTTTCATCCAAATTTGAAAGAAATTAAT 436
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  :

```

```

138 uasn.....ArgTyrAsnIleValAlaSPargAsnAsnA 149
437 ATTAGCGTCGACATTCACACCCCTTACAAAGCGGATTANCATGGCGGT 486
149 laPro.....SerLeuAspPheHisAlaProArg 158
487 TTGCATTAATTTGTGCAGATGCAGAAACCTGTGAAATGACGAGTAC.. 534
159 LeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaGlnI 175
535 .ATGAGGGGAATFACCTATTCGATTAAGAAAAATATCCCGAGCTGTCC 583
175 ValAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrProValPheTyr 192
584 GCATCGCTCAGGACACACACTATGCGCTATGATGATGATGACAAACGCG 633
192 rGlnGlySerGlyThrGlnTyrIleLys.....AspSerAsnGly 205
634 GATTATCTCTACTCCGCG.....GCATGTTAATTGGCGGCAATAC 674
206 GlnLeuThrGlnMetGlyGlyAlaTyrSerTrpLeuThrGlyGlyTyrVa 222
675 ACATATGCGAGGTGGGGAATAATAGCC.....GTANTACTTTGAGCG 718
222 LglSerLeuSerSerTyrGlnAsnGlyGlnMetIleSerTrpSerSerg 239
719 GCGATGTGCGCCATGCCAAGACTATGCGCTATGCGCATGTCAGATGCG 768
239 LysLeuValPheAspTyrLysLeuAsnGlyAlaMetProIleTyrGlyGlu 255
769 GCAGGCGACAGCGGTTCCGCAATGTTTATTATGACAAACAAACAATTA 818
256 AlaGlyAspSerGlySerProLeuPheAlaPheAspTrpValGlnAsnLys 272
819 ATGCGTGCACAGCGAGTTTACAAACGCGCTACCCCTTATCCGCGAGG 868
272 strPValLeuValGlyValLeuThrAlaGlyAsnGlyAlaGlyGlyAyg 289
869 AAAACGGTTTCCAGCTGATACGCAAGATGGTTCTACGATGACATTTAC 918
289 LysAsnStrpAlaValIleProLeuAspPheIleGlyGlnLysPheAsn 305
919 AGAGGGGATACACATACCGCTNTTTTGAACCGCGAGTACGAGACATTT 968
306 GluAspAsnAspAlaProValThrPheArgTrpSerGlyGlyAlaLe 322
969 TTCCCTTACATCCACAAACAAC...GGTACGGGTACGTAACAGAAACCA 1015
322 uGluTrpSerPheAsnSerSerThrGlyAlaGlyAlaLeuThrGlnGlyT 339
1016 ACGAAAAAGTNTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGACTGTT 1065
339 hr..... 339
1066 GAGCATCTTTGATGAACACTGATAAGAACGATTACGGCGAGGGG 1115
340 .....ThrThrTyrAlaMetHisGln 346
1116 TGTTAATCAGTACCGTCCAAAGTTAAACAAGGTGAAAACTTTCTTTTA 1165
346 Y.....GlnGlnGlyAsnAspLeuAsnAlaGlyLysAsnLeuIlePhe. 360
1166 TCGATTACGGCAACGCAAACTCATCTTATCAAAACAATCAACCAAGCG 1215
361 ..GlnGlyGlnAsnGlyGlnIleAsnLeuLysAspSerValSerGlnGly 376
1216 GCGGCGGGGTTGTATTGATTTGAGGTGATTTAGCGTCTCCCGGAACAA 1265
377 AlaGlySerLeuThrPheArgAspAsnTyrThrValThrThrSerAsnGln 393
1266 CGAAGCTGCGCAAGCGCGGCTTCATCATGACGTAAAGACAGTACCGTTA 1315
393 ySerThrTrpThrGlyAlaGlyIleValValAlaSPargAsnGlyValSerVal 410
1316 CTTGAAAGCTAAACGGCGCTGCAAAACGCGCCCTGTGTCCAAATCGCAAA 1365
410 snTrpGlnValAsnGlyValLysGlyLysAsnLeuHisLysIleGlyGlu 426
1366 GGCACGCTGCACGTTCAAGCCAAAGGGGAAACCAAGGCTCGATCAGCT 1415
427 GlyThrLeuThrValAlaGlnGlyThrGlyIleAsnGlnGlyGlyLeuLysVa 443
1416 GCGCGACGGTACAGTCAATTTGGATCGACGCGACGACGATTAAGGCAAA 1465
443 LglYAspGlyLysValValLeuAsnGlnGlnAlaAspAsnLysGlyGlnV 460
1466 AACAGCCTTTAGTGAATCGCCTTGTCAGCGCGACGCGGTACGCTGCAA 1515
460 alGlnAlaPheSerSerValAsnIleAlaSerGlyArgProThrValVal 476
1516 CTGAATCCGATATACAGTTCACACCCCGCAACACTATTTCCGGCTTTC 1565
477 LeuThrAspGluArgGlnValAsnProAspTrpValSerTrpLysTyrAr 493
1566 CGCGGACGTTTGGATTAAACGGGACATTCGCTTTCCTCCACCGTATTC 1615
493 gGlyGlyThrLeuAspValAsnGlyAsnSerLeuThrPheHisGlnLeu 510
1616 AAAATACCGATGAAGGGCGCATGATTGNCNATTAATGCCACACACACA 1665
510 yAlaAlaAspTyrGlyAlaValLeu...AlaAsnAsnValAlaSParg 525
1666 TCCACGCTTACCATTCACGGAATGAAGTATTAACACACCGAGTGTAA 1715
526 AlaThrIleThr..... 529
1716 GAATATCATATGACTTAATTAC.....AGCAAGAAATTTGCCCTACA 1756
530 .....LeuAspTyrAlaLeuArgAlaAspLysValAlaLeuAla 542
1757 ACGGTTGG..... 1764
542 snGlyTrpSerGlnSerGlyLysGlyThrAlaGlyAsnLeuTyrLysTyr 558
1765 .....TTTGGCGAAGAAAGATACGAC 1784
559 AsnAsnProTyrTrpAsnThrThrAspTyrPheIleLeuLysGlnSerTh 575
1785 C..... 1785
575 rTyrGlyTyrPheProThrAspGlnSerSerAsnAlaThrTrpLuphev 592
1785 ..... 1785
592 alGlyHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsn 608
1786 .....AAAACGACGCGCGCTCAACCTTGT 1811
609 ThrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnValAs 625
1812 TTACCAG...CCGCGCGCAGAGACCGCACCCGCTCTTCCGCGCGAA 1858
625 PAsnArgLeuProGlnGlyValThrGlyAlaLeuValMetAspLysAla 642
1859 CAATTTTAACGCAACATCCGCAACAAAGGCAAACTGTTTTCAGC 1908
642 LaAspIleSerGlyThrPheThrGlnGlnLysAsnGlyArgLeuThrLeuGln 638
1909 GCGAGACGCGACCGCGCTTACATCAT..... 1938
659 GlyHisProValIleHisAlaTyrAsnThrGlnSerValAlaAspLysLe 675
1939 .TTAGGAACGCGGTGTCAAAAATGGAAGTATCCCAAGAGAGAAATCG 1987
675 uAlaAlaSerGlyAspHisSerVal.....LeuThrGlnProThrSerp 690

```



```

1988 TGTGGCAACAGCACTGATCMACCGACGTTTAAAGCGAATAATTCAT 2037
      ::::::::::::::::::::
690 heserIngluAspTrrgluAsnArgSerThrPheAspArgLeuSer 706
      ::::::::::::::::::::
2038 ATTGAGGGCGGCGAGGGGTGATTCCCGCAATGTT.....GCCAA 2078
      ::::::::::::::::::::
707 LeuIysAsnThrAspPheGlyLeuGlyArgAsnAlaThrLeuAsnThr 723
      ::::::::::::::::::::
2079 AGTGAAGGCGGATTCGCAATTTGAGCAATGACGCCAAGCATTTTGTG 2128
      ::::::::::::::::::::
723 rIleGlnAlaAsp.....AsnSerSerValThrLeuGly 734
      ::::::::::::::::::::
2129 TCCGACCCGATCAAGCCATACATGTGACAGCTGGAGCTGGAAGCT 2178
      ::::::::::::::::::::
735 .....AspSerArgValPheIleAspIysAsnAspGlyGlnGly 747
      ::::::::::::::::::::
2179 CTGCAAAATGTGTGCA.....NAANCAATTACCGACGATTAAGT 2219
      ::::::::::::::::::::
748 ThrAlaPheThrLeuGluGluGlyThrSerValAlaThrLysAsp.... 762
      ::::::::::::::::::::
2220 GATTGCTTCATGACTAGACAGCAGTACGCGCAGTGTAGCTGAGCT 2269
      ::::::::::::::::::::
763 .....AlaAspIysSerValPheAsnGlyThrValAsnLeuAsp 776
      ::::::::::::::::::::
2270 ATTAGCANTNTTAAANCTCNCNGGCGCTGNNCACTTAANGCAATCT 2319
      ::::::::::::::::::::
776 snGlnSerValLeuAsnIleAsnAsp.....IlePheAsnGlyGlyLe 790
      ::::::::::::::::::::
2320 AGTGCAAATGGCGATACAGCTATATACAGTACGCCACACAGCCACCA 2369
      ::::::::::::::::::::
791 GlnAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804
      ::::::::::::::::::::
2370 CGGCAACTTACGCTGTGGCAATGCCACAGCAATTTAATCAAGCA 2419
      ::::::::::::::::::::
805 .....ValleuGlyAsnSerThrLeuThrSerThrAlaLeu 817
      ::::::::::::::::::::
2420 CATTAAGCGCAACNCATCGGNTTGGCGCAATGCTTCAATT..... 2460
      ::::::::::::::::::::
817 snLeuAsnGlyGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAsp 833
      ::::::::::::::::::::
2461 .....AATCTAAGCAACACGCCGCCACAA..... 2484
      ::::::::::::::::::::
834 GlyProValAsnIleSerAspAlaAlaLeuSerLeuAsnSerArgProAs 850
      ::::::::::::::::::::
2485 .....AACGCAAGTCTGA 2497
      ::::::::::::::::::::
850 pGluValSerHisThrLeuProValTyrAspTyrAlaGlySerTrr 867
      ::::::::::::::::::::
2498 CGCTT...TCGACAAAGCTAAGCAACGTAAGC...CATTCGCGACTC 2541
      ::::::::::::::::::::
867 snLeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeu 883
      ::::::::::::::::::::
2542 AACGGCAATGTCTCCCTAGCCGATAG..... 2568
      ::::::::::::::::::::
884 SerGlyAsnIleAsnValGlnAspIysGlyThrValThrLeuGlyGly 900
      ::::::::::::::::::::
2569 .....GCAG 2572
      ::::::::::::::::::::
900 uGlyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSer 917
      ::::::::::::::::::::
2573 TATTCATTTTGAACACAGCCGCTTTACCGACAACCTCAGCGGAGCAAG 2622
      ::::::::::::::::::::
917 eupheAsnGlyTyrArgAsnIleTrrPserGlySerLeuAsnAlaProAsp 933
      ::::::::::::::::::::
2623 GANACAGCATTTACACTTAAAGACAGCAATGACGCTCGCGTACAGCAC 2672
      ::::::::::::::::::::
934 AlaThr...ValSerMetThrAspThrGlnTrrPserMetAsnGlyAsn 949
      ::::::::::::::::::::
2673 GCAATTAAGCAATTAACCTTGACAAAGCCACCATTAACACTCAATCCG 2722
      ::::::::::::::::::::
949 rThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsnGly 966
      ::::::::::::::::::::
2723 CCAATCCGCAAGATGTGACAGCGCGCAACCGGACAGNGTGTACAGACG 2772
      ::::::::::::::::::::

```

```

966 Ly..... 966
2773 CCGCGCCGCCGCTGCGCGCGCTTCCCTATTATCCGTTACACGCCCACTTC 2822
      ::::::::::::::::::::
966 ..... 966
2823 GSTAATATCCGTTTCAACAGCTGACGTTAAGCGCAATTAAGACGTC 2872
      ::::::::::::::::::::
967 ....ThrsrProPheThrThrLeuThrThrAspAsnLeuAspAlaValG 982
      ::::::::::::::::::::
2873 AAGAAATTCGCGCTTATGTCGAACTCTCGGCTACCGGAGGAGCAAA 2922
      ::::::::::::::::::::
982 InserAlaPheValMetArgThrAspLeu.....AsnLysAlaAspLys 996
      ::::::::::::::::::::
2923 TTGAAGCTGGGCGAAAGTTCCGAAGGAACTTACACTTGGCGGTCAACA 2972
      ::::::::::::::::::::
997 LeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTrrValAs 1013
      ::::::::::::::::::::
2973 TACCGGCAACGACCCGTAACCTCGATCAATTG..... 3006
      ::::::::::::::::::::
1013 nPheLeuLysLysProSerAsnLysAspThrLeuAspIleProLeuValS 1030
      ::::::::::::::::::::
3007 .....ACGTA 3012
      ::::::::::::::::::::
1030 eAlaIleProGluAlaThrAlaAspAsnLeuPheAlaSerThrArgVal 1046
      ::::::::::::::::::::
3013 GTGGAAGGGAAGAACACAAACCGCTGTCCGAAACCTTAATTTACACCT 3062
      ::::::::::::::::::::
1047 ValGlyPheSerAspValThrProIleLeuSerValAlaArgLysGluAsp 1063
      ::::::::::::::::::::
3063 GCAAAAGCAACAGTGCATGCCGCGCGGTGCGCTTACCACTATCCGCA 3112
      ::::::::::::::::::::
1063 LysLysGlyLeuTrrValLeuAspGly.....TyrGlnValAlaArg 1077
      ::::::::::::::::::::
3113 AAGAGCGCGAATTCGCGCTCATATCCGGTCAAGAAAGAGCTTTC 3162
      ::::::::::::::::::::
1077 snAspArgLysGln..... 1080
      ::::::::::::::::::::
3163 GACAAACTCGCGCAAGCAGAACCCAAACAGCGGAGAAAGAACAC.. 3210
      ::::::::::::::::::::
1081 .....GlyLysAlaAlaIleThrPheMetHisIleSerTyrAsnAs 1094
      ::::::::::::::::::::
3211 .....GCCAAAGCCTTGACGCGCGCTGATTCGGCGCGCGCGCG 3247
      ::::::::::::::::::::
1094 nPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeuArg 1111
      ::::::::::::::::::::
3248 ATGCCCGCGAAGAACAGAAAGCGTTGCCGAAACCGCGCGCGAGGCG 3297
      ::::::::::::::::::::
1111 splLeaGlnGlyAlaGlyThrTrrValArgLeuLeuAsnGlnGlySerGly 1127
      ::::::::::::::::::::
3298 GCGGAAATATGTCGC.....ATTATGACGCGGAGGA 3329
      ::::::::::::::::::::
1128 SerAlaAspIysGlyThrPheThrAspHisTyrThrLeuLeuGlnMetGlyAl 1144
      ::::::::::::::::::::
3330 AGAGAAAAAAGCGGTGACGCGGATTAAGAC..... 3360
      ::::::::::::::::::::
1144 aaSprArgLysHisGlnLeuGlnSerMetAspLeuPheThrGlyValMet 1161
      ::::::::::::::::::::
3361 .....AAGCCTTGCGCAACAGCGCGAAGCGGAGGAAAC 3393
      ::::::::::::::::::::
1161 LaThrTyrThrAspThrAspAlaSerAlaAspLeuTyrSerGlyLysThr 1177
      ::::::::::::::::::::
3394 CGCGCGGNTACACAGCGCTTCCCGCGCGCGCGGCGCGCGGATTT 3443
      ::::::::::::::::::::
1178 LysSerTrrPglLysGlyLysPheTrrAlaSerGlyLeuPheArgSerGlyAl 1194
      ::::::::::::::::::::
3444 GCGGCAACGCGACGCCCAACCGCAACCTCAACCCCAAGCGGAGCGAC 3493
      ::::::::::::::::::::
1194 aTyrPheAspValIleAlaLysTrrIleHisAsnGlnLysLysTyrAsp 1211
      ::::::::::::::::::::
3494 TGATNAGCCGTTATGCCAATACCGGTTTGAATTTCCCGCACGCTC 3543
      ::::::::::::::::::::

```

1211 eu.....AsnPhenAlaGlyAlaGlyLysGlnAsnPhenArgSer..... 1223
 3544 AACAGCGTTTCCGCTACAGAGAA.....TTGGA 3575
 1224 HisSerLeuThrAlaGlyAlaGlyLysGlnAsnPhenArgSer..... 1240
 3576 CCGCGGTGGTCCGAGACCGCCGCAACGCGTTGGACACAGCAGCATCC 3625
 1240 pThrThrPheValGluProGlnAlaGluLeuValTrpGlyArgLeuGln 1257
 3626 GGNACACCAACTACCGTTCCGACAGATTTCGCGCTACCGCAACAA 3675
 1257 LysGlnThrPheAsnTrpAsnAspSerGlyMet..... 1267
 3676 ACCGACCTGCGCAATCGGTATGCAAGAAAACCTCGCAGC..... 3717
 1268AspValSerMetArgArgAsnSerValAlaAsnProLeuVal 1280
 3718 .GGCGCGCTGGCAGTCCTGTTTTCGACACACGAGCAACAAACANCTTCG 3766
 1280 LysArgThrGlyValVal.....SerGlyLysThrPheSerGlyLysA 1295
 3767 ACAGAGCGCATCGCAACTCGGACGCTTTCGCGCGCGCTTTCGCG 3816
 1295 sPTPserLeuThrAlaArgAlaGlyLeuHisTyr..... 1306
 3817 CAATACGCGCATCGGAGGTTCGACATCGCATCGACAGCGCGCGGCTT 3866
 1307GluPheAspLeuThrAspSerAlaAspValHisLe 1318
 3867 TAGCAGCGCATCTTCAGACGCGCATCGAGCAAAATCCGCGCGCG 3916
 1318 ulysAspAlaAlaGlyGlnHisGlnHisLeuValArgLysAspSerArgm 1335
 3917 TGCTGCAAT...TAGCGCATTCAGCAGCATAC..... 3945
 1335 eLeuTyrGlyValGlyLeuAsnAlaArgPheGlyAspAsnThrArgLeu 1351
 3946CGCGCGGTTCGCGGATTCGCGCATCGCAACGCTA 3980
 1352 GlyLeuGluValGluArgSerAlaPheGlyLysTyrAsnThrAspAspAl 1368
 3981 CATCGCGCGCAACG...CGCTAT 3999
 1368 alLeuAsnAlaAsnLeuArgTyr 1375
 seq_name: p1r2:S57664
 seq_documentation_block:
 IGA-specific metalloendopeptidase homolog sepa precursor - Shigella flexneri
 C:Species: Shigella flexneri
 C:Date: 19-Oct-1995 #sequence.revision 03-Nov-1995 #text_change 08-Dec-2000
 C:Accession: S57664; S69769; S69768
 R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsot, C.
 submitted to the EMBL Data Library, February 1995
 A:Description: Characterization of sepa, the major extracellular protein of Shigella flexneri
 A:Reference number: S57664
 A:Accession: S57664
 A:Molecule type: DNA
 A:Residues: 1-1366 <BEN>
 A:CROSS-references: EMBL:Z48219; NID:9886952; PIDN:CAA8252.1; PID:9886953
 R:Benjelloun-Toulimi, Z.; Sansonetti, P.J.; Parsot, C.
 Mol. Microbiol. 17, 123-135, 1995
 A:Title: Sepe, the major extracellular protein of Shigella flexneri: autonomous secreted
 A:Reference number: S69768; MUID:96020667
 A:Accession: S69769
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-66:1077-1366 <BEN>
 A:CROSS-references: EMBL:Z48219; NID:9886952; PIDN:CAA8252.1; PID:9886953
 A:Experimental source: strain M90T (serotype 5)
 A:Accession: S69768
 A:Molecule type: protein

A:Residues: 57-61,'0',63-72;'X',544-557;'X',1057-1068 <BBF>
 A:Note: 6-Glu, 14-Thr, 18-Gly and 26-Gln were also found
 C:Gene: sepa
 A:Gene: sepa
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: extracellular protein
 F:1-56/Domains: signal sequence #status predicted <SIG>
 F:57-1366/Product: IGA-specific metalloendopeptidase homolog sepa #status predicted <
 alignment_scores:
 Quality: 1003.00 Length: 1561
 Ratio: 1.263 Gaps: 60
 Percent Similarity: 50.865 Percent Identity: 25.881
 alignment_block:
 US-09-303-518d-651 x S57664 ..
 Align seg 1/1 to: S57664 from: 1 to: 1366
 22 ACACCGAAGAACACCGCAAGCCCGAAGACCGCGCGCTGCTC 71
 27 ThrCysLysSerHisArg.....ArgLeuSe 35
 72 GCCTGCTTACTTAGCCATATGCTGTGCTGGCATTTCCCAAGCTT 121
 35 rArgArgValIleLeuThrSerValAlaLeuSerLeuSerSerAlat 52
 122 GG.....CGCGACACACTTATTTCGCGCATCGACCAATATCAT 162
 52 rProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68
 163 CGCGACTTTGCGCAAAATTAAGGCAAGTTGCAAGTGGCGGCGCAAGATAT 212
 69 ArgAspPheAlaGlnLysGlnGlnPheThrProGlyThrThrAsnI 85
 213 TGAAGTTTACACAAAGAGGAGTTGTCGCGCAATATGACACAAG 262
 85 eSerIleTyrAspLysGlnGlnLysAsnLeuValGlyLys..LeuAspLysA 101
 263 CCGCGATGATGATTTTCTGTGTGTCGCGGTAC..... 297
 101 lProMetAlaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117
 298 ...GGCGTGGCGCATTTGTCGCGCATATATATTTGAGCGTGGCACA 344
 118 ProGlyAspHisThrLeuTyrSerProGlnTyrValIleThrAlaLysH 134
 345 TAACGCGCGCTATACACAGTTCATTTGGTCGCGAAGCAAGAAATCCG 394
 134 sValSerGlySerAspThrMetSerPheGly..... 144
 395 ATCAGCACCGTTTTTCTTACCAATTTGTGAAGAAATATATATTAAGCT 444
 145TyrAlaLysAsnThrTyrThrAla 152
 445 GACATTCACACCTTACAAGGC...GATTACATATGCGCGTTTGA 491
 153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
 492 TAAATTTGTCACAGATCGACAACTGCGCAAGAGACAGACATGAGG. 540
 169 rLysLeuValThrIleValAlaProAlaGluVal...SerAspIleLys 185
 541GGAAATACCTATTCGATTAAGAAATAATCCGAGCGTGC 585
 185 lValIleSerGlyAlaTyrGlnAlaGlyLysArgPheThrArgLysPheTyrArg 201
 586 ATCGGCTCAGACACACACTATTTGGCTTATGATGATGACAAACGGCGA 635
 202 LeuLysGlyGlyMetGlnTyrValLys.....AspLysAsnGlyAs 215
 636 TTTATCTAC.....TCGCGCGCATGTTAATTCGCGCGCAATACAC 676

[illegible]


```

1211 erSerHisSerTrpTyrAlaGlyAlaGlyValGlyTyrArgTyrHisLeu 1227
      ::::: ||| ::::: |||||
4027 .....GAAACGTCATATCGCCAC 4046
1228 ThrLysGluSerTrpValGluProGlnIleGluLeuValTyrGlySerVal 1244
      : ||| |||||
4047 CCCCCTCTTTCGTCACCGTACCGGCGGCGCATTAAGCGGATTAAT 4096
      : ||| |||||
1244 LserGlyLysAlaPheSerTrpGluAlaArgGlyMetAla.....LeuS 1259
4097 CATTCAACCGCGCCACACATNTCCATGCACNCCTTATTNAGCCGTCC 4146
      ||::|||
1259 erMetLysAspLysAspTyrAsnProLeuIleGlyArgThrGlyValAsp 1275
4147 TATACGATGCGCGCTCGGCG.....AAAGTCGGAACAGC..... 4182
      ||| |||||
1276 ValGlyArgAlaPheSerGlyAspAspTrpTyrSileThrAlaArgAlaG 1292
4183 .....GTCAATACCGCGCTATTGCGTCAGC 4207
      ::||| ::::: ||||
1292 LLeuGlyTyrGlnPheAspLeuLeuAlaAsnGlyGluThrValLeuGlnA 1309
4208 ATTTCGGCAAAACCCCGCAGTCGGAATGG..... 4236
      ||::: |||
1309 sPalasSerGlyLysArgPheGlyGlyLysAspSerArgMetLeu 1325
4237 .....GGCGTAACGCGGAATCAAA 4257
      ||::: |||||
1326 MetThrValGlyMetAsnAlaGluIleLys 1335
seq_name: p1r2:S61319

```

```

seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C:Species: Neisseria meningitidis
A:Variety: NG117
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61319
R:Lombolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseri
A:Reference number: S61314; MUID:95302961
A:Accession: S61319
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-508 <LOM>
A:Cross-references: EMBL:X82483; NID:9732881; PIDN:CA57866.1; PID:9732882
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
  Quality: 870.00      Length: 553
  Ratio: 2.451        Gaps: 17
  Percent Similarity: 64.195      Percent Identity: 36.890

```

```

alignment_block:
US-09-303-518D-651 x S61319 ..

```

```

Align seg 1/1 to: S61319 from: 1 to: 508

```

```

202 GCGAAGATATTGAGTTCACACAAAAGGAGATTGGTCGCGCAATC 251
      ||| |||||
1 AlathrAspLeuSerValLysAsnGlnGlnAsnIleGlyAsnAl 17
252 AATGACAAGACCCGATGATTGATTTTCTGTGTCGCGGTAC...G 298
      ::::: |||||
17 aleuSerAsnValProMetIleAspPheSerValAlaAspAlaAsnLysA 34
299 GCGTGGCGCATTTGGTGGCGCATATATATTGTGAGCTGGACAT... 345
      ::::: |||||
34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

```

```

346 .....AACGGCGCTATAACAGTTGATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlnHisAsnAs 67
372 TGTGCGGAGAGAGAAAGNATCCCGATCGACACCGCTTTCTTACCAATTG 421
      |||||
67 pValAlaAspLysGluAsn.....GluTyrArgValV 78
422 TGAAGAAGAAATATATAGCTTCGACATTCACACCTTACACAGCG... 468
      ||::: |||||
78 alGluGlnAsnAsnTyrGluProHisLysAlaTrpGlyAlaSerAsnLeu 94
469 .....GATTNCAATATGCGCGCTTTCGATAATTGTTCACAGA 506
      ||| ::||| ||::: |||||
95 GlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLysPheValThrGl 111
507 TGCAGAACCTGTCGGAATGACGAGTGCATGAGGCG...AATACCTATT 553
      ::::: |||||
111 uValAlaTrpIleAlaProThrAspAlaGlyGlyLeuAspThrTyrL 128
554 CCGATTAAGAAAATATCCCGAGCGGTCCGATCGGCTGACAGACACAC 603
      ::||| :::::
128 ysAspLysAsnArgPheSerSerPheValArgValGlyAlaGlyArgGln 144
604 TAT.....TGCCGTTATGATGATGACAAACAGGC... 633
145 LeuValTyrGlyLysGlyAlaTyrHisGlnGluLysGlnLysGlyTyr 161
634 .....GATTATCTTACTCCGCGCATGTTAATGGGCGC... 669
161 rAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleAlaIleThr 178
670 .....AATACACAT...ATGCAG 684
178 roTyrLysAspLysAlaAsnIleAspGlnThrMetAsnThrGlnGlyLeu 194
685 GGTGGGGAATTAATGCGGTANTAGTTGACGCGGCATGCGCCATGC 734
      |||||
195 GlyPheGlyAsnHisAsnThrHisTyrSerAlaGluLysGlnGln 211
735 CAACGACTATGCGCCATGCGGATTCGAGTGGCGGACGACAGCGTT 784
      :::::
211 aleuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyAspSerGlyS 228
785 CGCCATGTTTATTATGACAAAACAAATAAATGCTGCTCAACGGA 834
      |||||
228 erProLeuPheAlaPheAspLysGlnLysAsnGlnTyrValPheLeuGly 244
835 GTTTACAAACCGGCTACCTTATTCGCGGACGAGAAACGTTTCCAGCT 884
      :::::
245 ThrTyrAspTyrTrpAlaGlyTyrGly..... 253
885 GATACGCAAGATTTGTTCTACGATGACATTTACAGAGCGATACACATA 934
      ::|||
254 ....LysLysSerTrpGlnGluTrpAsnIleTyrLysLysGlu..... 266
935 CCGTCTTTTGAACGCGCGCATACGGACATTTTCTTACATCAAC 984
      |||
267 .....PheAlaAspLys.....IleLysGln 273
985 AACACGCTACGCGGTACGTAACAGAAACGAAACG..... 1023
      :::::
274 ArgAspAsnAlaGlyThrIleLysGlyAsnArgGlnHisHisTrpAsnI 290
1024 .....GTTTCAATCCAAAGCTTAAGTACAGACAGTCCGACTGTTG 1066
      ::|||
290 ethrPheGlyThrAsnSerLysIleGlySerThrAlaValArgLeuAla 307
1067 ACGAATCTTGAATGAACGTAAAGAAACAGATTACCGCGGAGGGGT 1116
      ::|||
307 sn.....AsnGluArgAsp..... 311

```



```

885 GATACGCAAGATTGGTCTACGATCATTAC..... 918
254 ....LysLysSerTrpGlnGluTrpAsnIleTyrLysGluPheAla 269
919 .....AGAGCGATACACATACCGCTCTTTTGAACCGCCG 954
269 sPlsIleLysGlnArgAspAsnAlaGlyThrIle.....Lys 281
955 AGTAACGGA.....CATTTTCCTTACATCCACAAACAGGTACGGG 998
282 GLYAsnGlyGluHisIleTrpAsnIleThrPheGly..... 293
999 TACGTAACGAGAACCAACGAAAGGTTNCAATCCAAAGTTAAAGTAC 1048
294 .....ThrsAsnSerHisIleGlySer.....T 301
1049 AGCACTCCGACTGTTTGACGATCTTTGAATGAATGAATGAAGAACCA 1098
301 hAlaValAlaTrpGlu.....AlaGlyAsnGluArgAsp..... 311
1099 GTTTACGCGCAGAGGGGTGTATATACGTCACGTTAAACAGCG 1148
312 .....AlaAsnAsnGln 315
1149 TGAACACCTTCTTTATGATACGCAACGCAACACTCATCTTATCA 1198
315 yGlnAsnValThrPheGluAsp.....AsnGlyThrIleValLeuAspG 330
1199 ACAACATCAACGAGCGCGCGCTTGTATTTGAAGGTATTTTACG 1248
330 hAsnIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThr 346
1249 GTCTGCGCTGAAAACAACGAA...ACGTGCGCAAGCGCGCGCTCATAT 1295
347 ValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspVa 363
1296 CAGTGAACGACGTACCGCTTACTTGAAGTAACGCGCTGCAACGAC 1345
363 lAlaAspGlyLysLysValAlaTrpGlnValLysAsnProAsnGlyAspA 380
1346 GCCTGTCGCAAAATGGCAAGCGCGTCGACGTCACGTCACGCAAGG 1395
380 rGluAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVal 396
1396 AACCAAGGCTCGATCAGCGTGGCGACGTCAGTCATTTGGATCAGCA 1445
397 AsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLy 413
1446 GGCACACGATTAAGCAAAAAACAAGCCTTATGTAATCGGCTGTGNTCA 1495
413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430
1496 GCGGAGGGGTAGCGTGCACATGATGCCGATATCACTTACACCCGAC 1545
430 eGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 446
1546 AAACCTATTTTCGGCTTTCGGCGCGACGTTTGATTTAAACGGGACATC 1595
447 AsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAs 463
1596 GCTTCGTTCCACCGTATTAATAATACCGATGAAGGGCGATGTGNCN 1645
463 PheThrPheGlnHisIleArgAsnValAspGlyAlaArgGlyLeVala 480
1646 ATCATATTAACCAACACATCCACCGTTACATTAACGGAATGAAGT 1695
480 snHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerLeu 496
1696 ATTACACAACGAGTGTGAATATATCAAT 1725
497 lIleThrAspProLysThrIleSerIleHis 506

```

```

seq_name: p1r2:S61320
seq_documentation_block:
  Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (f
  C:Species: Neisseria meningitidis
  A:Variety: NGC65
  C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
  C:Accession: S61320
  R:Lomholt, H.; Poulsen, K.; Mogens, K.
  Mol. Microbiol. 15, 495-506, 1995
  A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neis
  A:Reference number: S61314; MIMD:95302961
  A:Accession: S61320
  A>Status: preliminary; nucleic acid sequence not shown
  A:Molecule type: DNA
  A:Residues: 1-508 <LOM>
  A:Cross-references: EMBL:X82484; NID:q732871; PIDN:CA57867.1; PID:q732872
  C:Superfamily: Iga-specific metalloendopeptidase
  C:Keywords: hydrolase; metalloprotease

```

```

alignment_scores:
  quality: 855.00      length: 559
  Ratio: 2.395        Gaps: 19
  Percent Similarity: 63.864      Percent Identity: 36.494

```

```

alignment_block:
  US-09-303-518D-651 x S61320 ..

```

```

Align seg 1/1 to: S61320 from: 1 to: 508

```

```

202 GCGAAGATATTGAGTNTACACAAAAAGGGAGTGTGGCGCAATC 251
1 AlaThrAspLeuSerValLysAsnLysGlnGlyAsnIleGlyAsnAl 17
252 AATGACAAAACCCCGATGATTTGATTTTCTGTGGTCCGTAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34
299 GCGTGGCGGCTTGTGGCGATCAATATATTTGACCGTGCACAT... 345
34 rGlyLeuAlaThrValAlaAspProGlnIleValAlaValLysValS 50
346 .....AACGGCGCATATACAGCTGATTT 371
51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67
372 TGGTGGGAAAGAAAGNAATCCGATCAGCACCGTTTCTTACCAATTG 421
67 pValAlaAspLysGluAsn.....GluTyrArgValY 78
422 TGAAGAAATAATATTATACCTGCACATTCACACCTTACAAAGGC... 468
78 aGlnGlnAsnAsnTyrGluProHisLysAlaTrpSerAlaSerAsnLeu 94
469 .....GATTNCAATATCCGCGTTCGATTAATTTGTCACGA 506
95 GlYArgLeuGluAspTyrAsnMetAlaArgPheAsnLysPheValThrG 111
507 TGCAGAACCTGTGGAATGACGAGTACATGAGAGGG...AATACCTATT 553
111 uValAlaProIleAlaProThrAspAlaGlyGlyGlyLeuAspThrTyrL 128
554 CCGATTAAGAAAATATCCGACGCGTCCGATCGGCTCAGGACACAC 603
128 yAspLysAsnArgPheSerPheValArgValGlyAlaGlyArgGln 144
604 TAT.....TGGCGTTATGATGATGACAAACACGGC... 633
145 LeuValTyrGlyLysGlyAlaTyrHisGlnGluGlyAsnGluLysGlyTy 161
634 .....GATTATCTTACTCCGCGCATGTTAATTGGCGGC... 669

```



```

84 rosnlnysasntrpshislsiglyasnngllyargleucluaspryrrasn 100
||:|||||:| ||| :||| ||| :||
||:|||||:| ||| :||| ||| :||
478 ATGCCCGCTTGCATTAATTTGTCACAGATGCAAGCAACCTGCGAATATGAC 527
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 MetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaProth 117
528 GAGTCAGCAGAGGGG...AATACCTATTCCGATTAAGAAAAATATCCGC 574
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 rserlaclaglyglyValGluThrTyrLysAspLysAsnArgPheSerG 134
575 AGCGTCTCCGATCGGCTCAGACACACTATTGCGTTATGATATGAC 624
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 lnphevalrvalglvalaglythrGlnPheGluTyrAsnSerArgTyr 150
625 AAACAGCGGATTTATCTTACTCTCCGCGCATTAATTTGGCGCAATAC 674
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
151 AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrPr 167
675 ACATATGACG.....G 685
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
167 ofyrGlnaspValAsnValThrSerAsnLeuasnngllyGluLeuIleG 184
686 GTTGGGAAATATATGCGGTANTTATTTGAGCGGCGATGCGCCATGCC 735
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
184 lypheglyaspasnserLysHisSerProGluLysLeuLysGluVal 200
736 AACGACTATGCCCTATGCCGATTCAGAGTGCGGCGACGCGAGCGGTC 785
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAspSerGlySe 217
786 GCCAATGTTTATTTATGACAAAACAACATTAATGCTGCTCAACGAG 835
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 rProLeuPheAlaTyrAspLysGlnGluLysArgTyrPalPheLeuGlyA 234
836 TTTTCA.....ACGCGTACCTTATTCGCGAGGAAAAACGCTTTC 879
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
234 lAtyrAspTyrTrpAlaGlyTyr.....GlnLysAsnSerTrp 246
880 CAG.....CTGATACGCAAGATGTTCTACGATGACATT...TACAG 920
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
247 GlnGluTrpAsnIleTyrLysGluPheAlaAspLysIleLysGlnAr 263
921 AGCGCATTCACATACCGTCTTTTGAACCGCGCAGTAACGGA.....C 964
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
263 gAspAsnIaGlyThrIle.....LysGlyAsnGlyLysHSH 276
965 ATTTTCTTACATCCCAACAACAGCGTACGCGTAAACGAAC 1014
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
276 lStrpAsnIleThrPheGly.....Thr 283
1015 AACGAAAGGTTNTCCATCAACAGCTTAAGTACAGACAGTCCGACGTT 1064
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
284 AsnSerLysIleGlySer.....ThrAlaValAlaGlyLeu.. 294
1065 TGACCAATCTTGAATGAACACTGATAAACAACAGTTTACGCGCAGGG 1114
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
295 .....AlaGlyAsnGlnLysAsp..... 300
1115 GTGTTATCAGTACCGTCCAAAGGTTAAACAACGTAACCTTCTTTT 1164
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 .....AlaAsnAsnGlnGlyAsnValThrPhe 309
1165 ATCGATTACGGCAACGGCAACCTCATCTTATCAACAACATCAACAG 1214
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 GluAsp.....AsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnG 324
1215 CGCGGCGCTTGTATTTTGAAGTGATTTACGCTCGCCTGAACAACA 1264
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
324 yAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyLysAsn 341
1265 ACGAA...ACGTGCAAGCGCGCGCTTCATCAGTGAAGAAGATGACC 1311
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||

```

```

341 snAspIleThrTrpLeucllyAlaGlyIleAspValThrAspGlyLys 357
1312 GTTACTTGGAAAGTAAACGGCGTGGCAACAGCCGCTGTCCAAATCG 1361
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 ValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleG 374
1362 CAAGGCGCGCTGCGCTTCACAGCCAAAGGGAACCAACAGCGCGCTCA 1411
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
374 yLysGlyThrLeucllyLeuAsnGlyThrGlyValAsnGlnGlyLeuL 391
1412 GCGTGGCGGACGTACAGTCATTTTGGATCAGCAGCGACGATTAAGGC 1461
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
391 yValIleGlyAspGlyThrValIleLeuAsnGlnGlnAlaAspAlaAsp 407
1462 AAAAAACAGCCTTTTACTGAATAGCCTTGTACAGCGGCGGCTACG 1511
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
408 LysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThr 424
1512 GCACTGAATGCGATATCATGTTCAACCGCGACCAACCTCTATTGGCT 1561
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 uValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheLys 441
1562 TTGCGGCGGACGTTTGATTTAAACGGGCAATTCGCTTGTCCACCGT 1611
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 heArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHis 457
1612 ATTCAAAATACCGTGAAGGCGCATGTTGNCATATATATGCCACAC 1661
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 lLeArgAsnValAspGlnGlyAlaArgIleValAsnHisAsnThrSerH 474
1662 AACATCCACCGTTCATTCAGGGAATGAAGATTAATACCAACAGGATG 1711
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 sAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAsnProAsn 491
1712 GTAAGAATATCAATAGACTT 1731
||:|||||:| ||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
491 erLeuSerValHisSerIle 497
seq_name: p1r2:S61331
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: NG093
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
R:Limholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neis
A:Reference number: S61314; MUID:95302961
A:Accession: S61331
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-507 <LON>
A:Cross-references: EMBL:X82482; NID:q732885; PIDN:CA457865.1; PID:q732886
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease
alignment_scores:
Quality: 847.50 Length: 548
Ratio: 2.394 Gaps: 15
Percent Similarity: 64.599 Percent Identity: 36.861
alignment_block:
US-09-303-518d-651 x S61331 ..
Align seg 1/1 to: S61331 from: 1 to: 507
202 GCGCAAGATATGAGTGTNTCAACAACAAAGGAGGACTGTGCGCAATC 251
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 AlatrAspLeuSerValLysAsnLysGlnGlyAsnIleGlyAsnAl 17

```

```

252 AATGACAAAGCCCGATGATGATTTTCTGTGGTGGCGGTAAC...G 298
    : : : : : : : : : : : : : : : : : : : : : : : :
17 aleuSerAsnValrPrometileAspPheSerValAlaAspValAsnLysa 34
    : : : : : : : : : : : : : : : : : : : : : : : :
239 GCGTGGGGCGCTGTGGCGCATATATATTGAGCGTGGGCAT... 345
    : : : : : : : : : : : : : : : : : : : : : : : :
34 rglleAlaThrValAlaAspProGlnTyrAlaValSerValLysHISAla 50
    : : : : : : : : : : : : : : : : : : : : : : : :
346 .....AACGGCGCATATAACACGTTGATTT 371
    : : : : : : : : : : : : : : : : : : : : : : : :
51 LysAlaGluValAlaHisThrPheTyrTyrGlnTyrAsnGlnHisAsnAs 67
    : : : : : : : : : : : : : : : : : : : : : : : :
372 TGGTGGGGAAGGAAGNAATCCCGATGACACCGTTTCTTACCAATG 421
    : : : : : : : : : : : : : : : : : : : : : : : :
67 pValAlaAspLysGluAsn.....GluTyrArzrValY 78
    : : : : : : : : : : : : : : : : : : : : : : : :
422 TGAAGAAGATATATTAAGCGCTGACAAATTCACACCGTTTACACGGC... 468
    : : : : : : : : : : : : : : : : : : : : : : : :
78 alGluGlnAsnAsnTyrGlnProHisLysAlaTrpGlyAlaSerAsnLeu 94
    : : : : : : : : : : : : : : : : : : : : : : : :
469 .....GATTACATATGCGCGTTTGCATTAATTTGCACAGA 506
    : : : : : : : : : : : : : : : : : : : : : : : :
95 GlyAlaGluGlnAspTyrAsnMetalAlaArpHeAsnLysPheValAlaTrpG 111
    : : : : : : : : : : : : : : : : : : : : : : : :
507 TGCAGAACCTGTGCAATGACGAGTGCACAGAGGGG...AATACCTATT 553
    : : : : : : : : : : : : : : : : : : : : : : : :
111 uValAlaProIleAlaProThrAspAlaGlyGlyLeuAspPheTyrTrL 128
    : : : : : : : : : : : : : : : : : : : : : : : :
554 CCGATTAAGAAATATCCGCGGTCGCGATCGGTCGCGACGACACAC 603
    : : : : : : : : : : : : : : : : : : : : : : : :
128 yAspLysAsnArpPheSerPheValArzrValGlyAlaGlyAlaArgGln 144
    : : : : : : : : : : : : : : : : : : : : : : : :
604 TAT.....TGGCGTTATGATGATGACAAACACGGC... 633
    : : : : : : : : : : : : : : : : : : : : : : : :
145 LeuValTyrGlnLysGlyAlaTyrHisGlnGlnLysGlnLysGlyTyr 161
    : : : : : : : : : : : : : : : : : : : : : : : :
634 .....GATTATCTTACTCCGGCGCATGTTATTTGGCGC... 669
    : : : : : : : : : : : : : : : : : : : : : : : :
161 rAspLeuArzrAspLeuSerGlnAlaTyrArzrValAlaIleAlaGlyThr 178
    : : : : : : : : : : : : : : : : : : : : : : : :
670 .....AATACACAT...ATGCAG 684
    : : : : : : : : : : : : : : : : : : : : : : : :
178 rOTyLysAspLLeAsnLLeAspGlnThrMetAsnThrGlnGlyLeuIle 194
    : : : : : : : : : : : : : : : : : : : : : : : :
685 GGTTCGGCAATTAATGCGGTANTTATGTTGACGGCGATGTCGCCATGC 734
    : : : : : : : : : : : : : : : : : : : : : : : :
195 GlyPheGlnAsnHisAsnLysGlnTyrSerAlaGlnGlnLeuLysGlnAl 211
    : : : : : : : : : : : : : : : : : : : : : : : :
735 CAACGACTATGGCCCTATGCCGATGCGAGTGGCGGACGACAGCGGTT 784
    : : : : : : : : : : : : : : : : : : : : : : : :
211 aleuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyAspSerGlyS 228
    : : : : : : : : : : : : : : : : : : : : : : : :
785 CGCCAAATGTTATTTATGACAAACAAACAAATTAATGTCGTCGTCACGA 834
    : : : : : : : : : : : : : : : : : : : : : : : :
228 erProLeuPheAlaPheAspLysGlnLysAsnGlnTTPValPheLeuGly 244
    : : : : : : : : : : : : : : : : : : : : : : : :
835 GTTTTACAAACCGCTACCTTATTCGCGGAGGAAACGGTTTCCAGCT 884
    : : : : : : : : : : : : : : : : : : : : : : : :
245 ThrTyrAspTyrTrpAlaGlyTyrGlnLysLysSerTrpGlnLysTrpAs 261
    : : : : : : : : : : : : : : : : : : : : : : : :
885 GATACGCAAGATGTTCTTACGATGACATTTACAGAGGCATACACATA 934
    : : : : : : : : : : : : : : : : : : : : : : : :
261 nIleTyrLysLysGlnPheAlaAspGlnIleLysGlnArzrAsp...Asn 277
    : : : : : : : : : : : : : : : : : : : : : : : :
935 CCGCTNTTTTGAACCGGACGTAACGACATTTTCTTCCATTATCCACAC 984
    : : : : : : : : : : : : : : : : : : : : : : : :
277 laGlyThrIleLysGlnTyrGlnGlnHisThrLysTrpLysThrThcGlyThr 293
    : : : : : : : : : : : : : : : : : : : : : : : :
985 AACACAGCGTACGCTACGTAACAGAAACCAACAAAGGTTTCCAAATCC 1034
    : : : : : : : : : : : : : : : : : : : : : : : :
294 AsnSerHisIleGlySerThr..... 300
    : : : : : : : : : : : : : : : : : : : : : : : :
1035 AAAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAAATCTTGAATGAA 1084

```

```

301 .....AlaValArzrLeuAlaAsn.....AsnGluAla 309
    : : : : : : : : : : : : : : : : : : : : : : : :
1085 CTGATTAAGAACCACTTATACGGCGAGGGGTGTTAATCAGTACCGTCA 1134
    : : : : : : : : : : : : : : : : : : : : : : : :
309 rGAsp..... 310
    : : : : : : : : : : : : : : : : : : : : : : : :
1135 AGGTTAAACACGCTGAAACCTTTCTTTATTCATTCAGGACGACGCA 1184
    : : : : : : : : : : : : : : : : : : : : : : : :
311 ...AlaAsnGlnGlnAsnValThrPhe.....GluAsnAsnGlyTh 324
    : : : : : : : : : : : : : : : : : : : : : : : :
1185 ACTCATCTTATCAACATCAACACAGGCGGCGGCGGTTGTTATTTG 1234
    : : : : : : : : : : : : : : : : : : : : : : : :
324 rLeuValLeuAspGlnAsnLLeAsnGlnGlnValAlaGlyLeuPhePhe 341
    : : : : : : : : : : : : : : : : : : : : : : : :
341 yseGlyAspTyrThrValLysGlyAlaAsnAsnAspIleThrTrpLeuGly 357
    : : : : : : : : : : : : : : : : : : : : : : : :
1282 GCGGCGCTTCATATCAGTGAACAGAGTACCGTTACTGTGAAGTAAACGG 1331
    : : : : : : : : : : : : : : : : : : : : : : : :
358 AlaGlyIleAspValAlaAspGlnLysLysValAlaTrpGlnValLysAs 374
    : : : : : : : : : : : : : : : : : : : : : : : :
1332 CGTGGCAACGACCGCGCTGCCAAATCGCAAGGACGACGTCAGTTC 1381
    : : : : : : : : : : : : : : : : : : : : : : : :
374 nProAsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGlnIleA 391
    : : : : : : : : : : : : : : : : : : : : : : : :
391 snGlyThrGlyValaAsnGlnGlnLysGlnLysValGlyAspGlyThrVal 407
    : : : : : : : : : : : : : : : : : : : : : : : :
1432 ATTTTGGATCAGCAGGACGACGATTAAGCAAAAAACCTTTAGTGA 1481
    : : : : : : : : : : : : : : : : : : : : : : : :
408 IleLeuAsnGlnLysAlaAspAlaAspLysLysValGlnAlaPheSerG 424
    : : : : : : : : : : : : : : : : : : : : : : : :
1482 AATCGGCTGTGNTCAGCGGCGAGGGGTACGGTCACTGAATGCCGATATC 1531
    : : : : : : : : : : : : : : : : : : : : : : : :
424 nValGlyIleValSerGlyArzrGlyThrLeuValLeuAsnSerSerAsnG 441
    : : : : : : : : : : : : : : : : : : : : : : : :
1532 AGTTCAACCCGCAACAATCTATTTGCGCTTTCGCGGCGAGCTTTGGAT 1581
    : : : : : : : : : : : : : : : : : : : : : : : :
441 nIleAsnProAspAsnLeuTyrPheGlyPheArzrGlyLysArzrLeuAsp 457
    : : : : : : : : : : : : : : : : : : : : : : : :
1582 TTAAACGGGATTCGCTTCGCGACGATTCGATTAACGACGATCAAG 1631
    : : : : : : : : : : : : : : : : : : : : : : : :
458 AlaAsnGlyAsnAspLeuThrPheGlnHisIleAlaAspValaAspGln 474
    : : : : : : : : : : : : : : : : : : : : : : : :
1632 GCGCATGATGNCNATCATTAATGCCAACAACATCCACGTTTACCATTA 1681
    : : : : : : : : : : : : : : : : : : : : : : : :
474 yAlaArzrGlyIleValAsnHisAsnThrGlyHisAlaSerThrIleThrLeu 491
    : : : : : : : : : : : : : : : : : : : : : : : :
1682 CAGGGAATGAAGTATTACACAACGAGTGTGAATATCAAT 1725
    : : : : : : : : : : : : : : : : : : : : : : : :
491 hrGlyLysSerLeuIleThrAspProLysThrIleSerIleHis 505
    : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: p1r2:S61330
seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (tr
C:Species: Neisseria meningitidis
A:Variety: SM1166
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61330
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61330
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-508 <LOW>
A:Cross-references: EMBL:X82486; NID:q732860; PIDN:CA57869.1; PID:q732861
C:Superfamily: Iga-specific metalloendopeptidase

```



```

376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyIleuLeuValGlyA 393
1421 ACGGTACACTCATTTGGATGACGAGCAGACGATTAAGGCAAAACAA 1470
393 spGlyThrValIleuAsnGlnGlnAlaAspAlaAspLysLysValGln 409
1471 GCCTTATAGTAATCGGCTGTGTCAGCGGACGAGGTTACGGTCAATGA 1520
410 AlaPheSerGlnValGlyIleValSerGlyArgGlyThrLeuValLeuAs 426
1521 TGGCCATATACGTTCAACCCGACAAACTCATTTCCGCTTGGCGCGC 1570
426 nSerSerGlnGlnIleAsnProAspAsnLeuThrPheGlyPheArgGlyG 443
1571 GACGTTTGATTTAAACGGGCAATCGCTTGTCCACCGCATTTCAAAAT 1620
443 LysArgLeuAspAlaAsnGlnLysAsnAspLeuThrPheGlnHisIleArgAsn 459
1621 ACCGATGAAGGGGCGATGATTCNCATCATATATGCCACACAACTCCAC 1670
460 ValAspGlnGlyAlaArgIleValAsnHisAsnThrAspArgAlaSerTh 476
1671 CGTTACCATTTACAGGCAATGGAAGATATTACACACCGAGTGTAAGATA 1720
476 rIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuSerV 493
1721 TC 1722
493 al 493
seq_name: p1r2:S61323
seq_documentation_block:
IGA-specific metalloendopeptidase (EC 3.4.24.13) - Neisseria meningitidis (strain HF48
N:Alternate names: IGA proteinase; immunoglobulin A1 proteinase
C:Species: Neisseria meningitidis
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61323; S61324
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria
A:Reference number: S61314; MUID:95302961
A:Accession: S61323
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82480; NID:g732875; PIDN:CA57863.1; PID:g732876
A:Experimental source: strain HF48
A:Accession: S61324
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOH>
A:Cross-references: EMBL:X82475; NID:g732879; PIDN:CA57858.1; PID:g732880
A:Experimental source: strain HF96
C:Genetics:
A:Gene: iga
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
Quality: 842.00 Length: 537
Ratio: 2.332 Gaps: 15
Percent Similarity: 67.225 Percent Identity: 37.058

```

alignment_block:

US-09-303-518D-651 x S61323 ..

Align seg 1/1 to: S61323 from: 1 to: 496

```

202 GCGAAGATATGAGTNTACACAAAGGAGTGTGCGCAATC 251
||| ||||| ||||| ||||| ||||| ||||| |||||
1 AlaThrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17

```

```

252 AATGACAAAGCCCGATGATGATTTTCTGTGGTCCCGTAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAlaAsnArg 34
299 GCGTGGCGGCAATGTGGCGGATCATATGATTTGAGCGAGTGCACATTAAC 348
34 rGlnThrLeuThrValIleAspProGlnIleThrAlaValSerValLysHisVal 50
349 GCGGCGTATTAACAAGTTGATTTTGTGCGGAAGAAAGMATCCCGAT... 396
51 LysGlnLysArgIleSerValLysGlnHisAsnGlnLysIleuAspVal 67
397 ...CAGCACCCTTTTCTTACCAATTTGAAAGAAATTAATTAAGC 442
67 lSerAsnAspLysGlnIleuThrArgSerValAlaGlnAsnAspLysGlnP 84
443 CTGACAAATTC...CACCCTTACAAAGCG...GATTNCAAT 477
84 roAsnLysAsnThrPheHisGlyAsnGlnGlnArgLeuGlnAspLysThrAsn 100
478 ATGCCGCGTTTGCATTAATTTGTACAGATGCAAGACCTGTGCAATGAC 527
101 MetaLalrLeuAsnLysPheValThrGluValAlaProIleAlaProTh 117
528 GAGTGACATGAGGGG...AATACCTATTCGCAATTAAGAAATTAATCCG 574
117 rSerAlaGlyGlyValGlyValGlyThrLysAspLysAsnArgPheSerG 134
575 AGCGTGTCCGATCGGCTCAGACACACACTATTCGCTTATGATGATGAC 624
134 lPheValArgValGlyAlaGlyThrGlnPheGluIleuThrLysAsnArgTyr 150
625 AAGCAGCGCATTTATCTCTCCGCGCATGCTTATTTGGCGCAATAC 674
151 AsnMetThrGluLeuSerArgAlaLysArgLysAlaIleAlaGlyThrPr 167
675 ACATATGCAG...G 685
167 oTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
686 GTTGGGGAATTAATGCGGTANTAGTTTGAAGCGGATGCGCCATGCC 735
184 lPheGlyAspAsnSerLysHisIleSerProGlnLysLeuLysGlnVal 200
736 AAGCATATGCGCCCTATGCGGATTTGACAGTGGCGGACGAGCGGCTTC 785
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuGlyAspSerGlySe 217
786 GCGAATGTTTATTTATGACAAACAAACAAATTAATGCTGCTCAACGAG 835
217 rProLeuPheAlaLysAspLysGlnGlnLysArgTyrValPheLeuGlyA 234
836 TTTTACAA...ACCGGCTACCTTATTCGCGGCAAGGAAACGGTTTC 879
234 lAspLysPheLysGlnGlnLysGlnLysGlnLysGlnLysGlnLysGln 246
880 CAG...CTGATPACGCAAGATGTTGTTCTPACGACATTTTACAGAG 923
247 GlnGluThrAsnIleLysLysGlnPheAlaAspLysGlnLysGlnLys 263
924 CGATACACATACCGTCTNTTTGAAACCGGCAATGACGAGCATTTTTCST 973
263 gaAsp...AsnAlaGlyThrIleLysGlyLysGlnHisIleThrLys 279
974 TTTACATCCAAACAAGAGGTACGGGTACGGTAAACAGAAACGAAAG 1023
279 hrThrGlyThrAsnSerHisIleGlySerThr... 289
1024 GTNTCAATCCAAAGCTTAAGTACAGACAGTCCGACTGTTGACGAAATC 1073
290 .....AlaValArgLeu..... 293

```

```

1074 TTGATGAACCTGATTAAGAACAGCTTACGGCAGGGGCTTAAATC 1123
      |||||
294 .....AlaGlyasn..... 296

1124 AGTACCGTCAAGGTTAAACAAGTGAACCTTCTTTATTCGATTAC 1173
      |||||
297 .....GluThrGlyAlaAsnAsnGlyGlnAsnValThrPhe.....Glu 309

1174 GGCACGCGCAACCTCATCTTATCAACAATCAACCAAGCGCGCGG 1223
      |||||
310 AsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGly 326

1224 TTTGATTGTAAGGTGATTATTCAGCTCCGCTGAAAACAAGCA...A 1270
      |||||
326 yLeuPhePheLysGlyAspPheThrValLysGlyIleAsnAsnAspIleT 343

1271 CGTGGCAAGCGCGCGGCTTCATTCAGTAAGACAGTACCGTTACTGG 1320
      |||||
343 hThrLeuGlyAlaGlyIleAspValAlaAspGlyLysLysValValTyr 359

1321 AAGGTAAAGCGCGTGCACCAAGCAGCCCTGTCCAAAATCGCAAGGAC 1370
      |||||
360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376

1371 GGTGACGTTCAAGCCAAAGGGGAAACCAAGCTCGATCAGCGGGCG 1420
      |||||
376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyA 393

1421 ACGGTACACTCATTTGGATCAGCAGGCGACGATTAAGCAAAACAA 1470
      |||||
393 spGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGlnLysValGln 409

1471 GCCTTAGGAATGCGCTTGTGTCAGCGGCGGATCGTCAAGTGA 1520
      |||||
410 AlaPheSerGlnValIleValIleValSerGlyArgGlyThrLeuValLeuAs 426

1521 TGCCGATTAATCAGTTCACCCGACAACTATTTGGCTTTCGCGAGC 1570
      |||||
426 nSerSerAsnGlnIleAsnProAspAsnLeuThrPheGlyPheArgGly 443

1571 GACGTTGATTTAAACGGGCGCATTCGCTTTCGTCACCGTATTCAAA 1620
      |||||
443 LysArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIleLeuArgAsn 459

1621 ACGGATGAAGGCGCGATGTCNCNATCATTAATGCACAAACAATCCAC 1670
      |||||
460 ValAspGlnGlyAlaArgIleValAsnHisAsnThrGlnHisThrSerTh 476

1671 CGTTACCATTAAGGAATGAATGATTACACAAACCGAGTGTAAAGATA 1720
      |||||
476 rIleThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerV 493

1721 TCATAGACTT 1731
      |||||
493 aHisSerIle 496

```

seq_name: p1r2:S61327

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm
C:Species: Neisseria meningitidis

A:Variety: MCG80

C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S61327

R:Lombolt, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neisseria

A:Accession: S61314; MUID:95302961

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <L0M>

A:Cross-references: EMBL:X82479; NID:9732883; PIDN:CA57862.1; PID:9732884

C:Superfamily: IgA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

alignment_scores:
Quality: 840.00 Length: 538
Ratio: 2.360 Gaps: 17
Percent Similarity: 66.171 Percent Identity: 37.732

alignment_block:
US-09-303-518D-651 x S61327 ..

Align seg 1/1 to: S61327 from: 1 to: 496

```

202 GCGAAGATTTGAGTNTACCAAAAAGGGAGTGGTGGCAATC 251
      ||| |||||
1  AlatrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
      |||||
252 AATGACAAAAGCCCGATGATGATTTCTGTGTGTCGCGTAAAC...G 298
      |||||
17  AluSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34
      |||||
299 GCGTGGCGCATTTGGTGGCGATCAATATATTGTGACCGTGCACATAC 348
      |||||
34  rGThrLeuThrValIleAspProGlnThrValAlaValSerValLysHisVal 50
      |||||
349 GCGGCGTATTAACAACGTTGATTTGTGCGGAAGGAAGNATCCGAT.. 396
      |||||
51  LysGlyAspGlyIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67
      |||||
397 ....CACACCGCTTTTCTTACCAAAATGTGAAGAAATATATTAAGC 442
      |||||
67  LSerAsnAspGlnAsnGlnTyrArgSerValAlaGlnAsnAspTyrGlnP 84
      |||||
443 CTGCAATTC....CACCTTACAAGCC.....GATTANCAT 477
      |||||
84  rAsnLysAsnThrPheHisGlyAsnGlnGlyArgLeuGluAspTyrAsn 100
      |||||
478 ATGCGCGCTTTGCAATATTTGTCACAGATCGACAACTGTGCAATGAC 527
      |||||
101  MetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaProTh 117
      |||||
528 GAGTGACATGAGGGG...AATACCTATTCGATTAAGAAAAATATCCG 574
      |||||
117  rSerAlaGlyGlyValGluThrTyrLysAspLysAsnArgPheSerG 134
      |||||
575 AGCGTGTCCGATCGCTCGACGACACCACTATGGCTTATGATGATGAC 624
      |||||
134  LuPheValArgValGlyAlaGlyThrGlnPheGluTyrAsnSerArgTyr 150
      |||||
625 AATACGCGCATTTATCTACTCCGCGCATGTGTAATGGCGGCAATAC 674
      |||||
151  AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrPr 167
      |||||
675 ACATATGCAG.....G 685
      |||||
167  oTyrGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
      |||||
686 GTTGGGAATATATGCGGTANTTAGTTGAGCGGCGATGTCGCGCATGCC 735
      |||||
184  LysPheGlyAspAsnSerLysHisHisSerProGluLysLeuLysGlyVal 200
      |||||
736 AATGACATGAGCCCTATGCCGATGACAGTGGCGGACGACACAGCGTTC 785
      |||||
201  LeuSerGlnAsnAlaLeuThrAsnTyrAlaValLeuLysSerGlySer 217
      |||||
786 GCGATGTTTATTATGACAAAACAACATTAATGGCTCTCTCAACGAG 835
      |||||
217  rProLeuPheAlaTyrAspLysGlnGluLysArgTyrValPheLeuLysA 234
      |||||
836 TTTTACAA.....ACCGGCTACCTTATTCGCGGAGGAAGAAAGTTTC 879
      |||||
234  lATyrAspTyrTyrPalaGlyTyr.....GlnLysAsnSerTyr 246

```



```

167 oYUGlnAspValAsnValThrSerAsnLeuAsnGlnGlyLeuIleG 184
686 GTTGGGAAATATGGCGTANTAGTTTGAGCGGCGATGCGCCATGCC 735
184 LYPheGlyAspAsnSerLysHisHisSerProGluLysLeuLysGluVal 200
736 AACGACTATGGCCCTATGCCGATTTGAGTGGCGGACGACGCGGTTC 785
201 LeuSerClnAsnAlaLeuThrAsnGlyAlaValLeuGlyAspSerGlySe 217
786 GCCAATGTTATTTATGCAAAACAACATTAATAGGCTGCTCAACGGAG 835
217 rProLeuPheAlaLysAspLysGlnGluLysArgThrValPheLeuGlyA 234
836 TTTTACAA.....ACCGCTACCTTTATTCGCGACGGGAAACGGTTTC 879
234 LArYAspLysTrpAlaGlyTrp.....GlnLysAsnSerTrp 246
880 CAG.....CTGATACGCAAAAGATTGGTTACGATGACATTTACAGAG 923
247 GlnGluTrpAsnIleLysLysGluPheAlaSerGluIleLysGlnAr 263
924 CGATACACATACCGCTCTTTTGAACGCGCAGTAAAGCAGCATTTTCTCT 973
263 GAsp...AsnAlaGlyThrIleLysGlyLysGlnHisHisTrpLysTr 279
974 TTACATCCACACACACGCGTACGCGTACGATACAGAACCCAGAAAG 1023
279 hrThrGlyThrAsnSerHisIleGlySerThr..... 289
1024 GTTTCATCCAAAGCTTAAAGTACAGACGCTCCGCTTTGACGATC 1073
290 .....AlaValArGlyLeu..... 293
1074 TTTGATCAAACTGATAAAGACAGATTACGCGGCGGCTGTATTC 1123
294 .....AlaGlyAsn..... 296
1124 AGTACCGTCCAAAGTTAAACACGCTGCAAAACCTTTCTTATGATTAC 1173
297 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe.....Glu 309
1174 GGCACGCGCAACATCTATTCACAAACATCAACACCAAGCGCGCGCG 1223
310 AsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyAlaGly 326
1224 TTTGATTTTGAAGTATTTTACGCTCCGCTGCAAAACCAAGAA...A 1270
326 YLeuPhePheLysGlyAspLysTrpValLysGlyIleAsnAsnSprIleT 343
1271 CGTGGCAAGCGCGCGCTTCATATCAGTAAAGACGATACCGTTACTTGG 1320
343 hrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysValValTrp 359
1321 AAAGTAAAGCGCGTGGCAACGCGCTGCAAAATCGCAAAAGCAC 1370
360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
1371 GCTGACGCTTCAAGCCAAAGGGAACCAAGCTGATACAGCGGGCGG 1420
376 rLeuGluIleAsnGlyThrGlyValAsnGlnGlyLysLeuLysValGlyA 393
1421 ACGGTACAGTCAATTTGATCAGACGACGACGATTAAGGCAAAACAA 1470
393 spGlyThrValIleLeuAsnGlnLysAlaAspSerAsnGlnLysValGln 409
1471 GCCTTTAGTAAATGGCTTGTACAGCGGACGAGGCTAGCGTCAACTGA 1520
410 AlaPheSerGlnValGlyIleValSerGlyArgArgThrLeuValLeuAs 426
1521 TGCCGATTCAGTTCAACCCGACAAACTATTTGCGCTTCGCGGCG 1570
426 nSerSerAsnGlnIleAsnProAsnLeuLysPheGlyPheAlaGlyG 443

```

```

1571 GACGTTTGATTTAAACGCGCATTCGCTTGTCCACCGGATTCAAAT 1620
443 LArGlyLeuAspAlaAsnGlnLysAsnLysPheTrpGluHisIleArgAsn 459
1621 ACCGATGAAGGCGCGATGATTCGNCATCATTAATGCAACACATCCAC 1670
460 ValAspGlnGlyAlaArgGlyLeuAlaAsnHisAsnThrGlyHisTrpSerTh 476
1671 CGTTACCATTCACGGAATGAATGATTTACACAAACCGAGTGGTAAGATA 1720
476 rIleThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerV 493
1721 TCATATGACTT 1731
493 ..... 496
493 aHisSerIle 496
seq_name: plr2:S61322

```

```

seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
C:Species: Neisseria meningitidis
A:Variety: HF54
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61322
R:Rimbold, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61322
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-496 <LOM>
A:Cross-references: EMBL:X82473; NID:g732877; PIDN:CAA57856.1; PID:g732878
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase, metalloproteinase

```

```

alignment_scores:
Quality: 815.00 Length: 534
Ratio: 2.309 Gaps: 15
Percent Similarity: 66.105 Percent Identity: 36.704

```

```

alignment_block:
US-09-303-518d-651 x S61322 ..

```

```

Align seg 1/1 to: S61322 from: 1 to: 496

```

```

202 GCGAAGATATTTGAGGTTTACACAAACAAAGCGAGTTGTCGCAATC 251
1 AlArThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 AATGACAAAGCCCGATGATGATTTTCTGTGCTGCGCGTAAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34
299 GCGTGGCGGATTTGGGCGGATATATGTCAGCGTGGACACATTAAC 348
34 rGluThrLeuThrValIleAspProGlnTrpAlaValSerValLysHisVal 50
349 GCGCGTATGAACAACGTTGATTTTGGTCCGAGGAAGAAATCCGAT.. 396
51 LysGlyAspSerGluIleSerGlyTrpGlyHisHisAsnGlyHisLeuAspVa 67
397 ....CACACCGCTTTTCTTACCAATTTGAAACAAATTAATTTAAGC 442
67 LSerAsnAspGlnAsnGlnLysArgSerValAlaGlnAsnAspLysTrp 84
443 CTGCAATTC.....CACCGTTACAAAGCG.....GATTANCA 477
84 rAsnLysAsnTrpHisHisGlyAsnGlnGlyArgLeuGluAspTrpAsn 100
478 ATGCCGCGTTTGCAATTAATTTGTCACAGATGCAGAACCTGCAATGAC 527

```



```

101 MetIaIArgLeuAsnLysPheValThrGluValAlaProIleAlaProTh 117
528 GAGTGCATGAGGGG...AATACCTATCCGATTAAGAAATAATATCCG 574
117 rseAlaIglYlGlyValGluThrGlyAspLysAsnArgPheSerG 134
575 ACCGTGCGCCATCGGCTCAGACACCACATATTGGCGTTATGATGATG 624
134 IupheValArgValGlyAlaGlyThrGlnPheGluThrIleAsnSerArgTyr 150
625 AAACACGGCGCATTTATCTACTCCGGCGCATGGTTAATGGCGCAATAC 674
151 AsnMetThrGluLeuSerArgAlaTyrArgTyrAlaIleAlaGlyThrPr 167
675 ACATATGAC...G 685
167 cYrGlnAspValAsnValThrSerAsnLeuAsnGlnGluGlyLeuIleG 184
686 GTTGGGGAATTAATGCGGTANTTAGTTCAGCGCGATGTGCGCCATGCC 735
184 IypheGlyAspAsnSerLysHisSerProGluLysLeuLysGluVal 200
736 AAGGACTATGGCCCTATGCCGATTCAGAGTGGCGGACGACAGCGGTTTC 785
201 LeuSerGlnAsnAlaLeuThrAsnTyrAlaValIleGlyAspSerGlySe 217
786 GCCAATGTTATTTATGACAAACAAACAATAATGCTGCACACGCGAG 835
217 rProLeuPheAlaTyrAspLysGlnGluLysArgTyrPheValPheLeuGlyA 234
836 TTTTACA...ACCGCTACCTTATTCGCGCAGGAGAAACGGTTTC 879
234 IaTyrAspTyrThrPalaGlyTyr...GlnLysAsnSerThr 246
880 CAG...CTGATACGCAAGATTGGTCTACGATGACATTTACAGAG 923
247 GlnGluThrAsnIleTyrLysGluPheAlaAspGluIleLysGlnAr 263
924 CGATACACATACCGTCTTTTGAACCGCGCAGTAAACGACATTTTTCCT 973
263 gAsp...AsnAlaGlyThrIleLysGlyTyrGluGlnHisThrPlyST 279
974 TTACATCCACAACAACGAGTACGGGTACGATACAGAAACCAACGAAAG 1023
279 hrThelYthrAsnSerHisIleGlySerThr... 289
1024 GTNTCAATCCAAAGCTTAAGTACAGACAGTCCGACTGTTGACGANTC 1073
290 .....AlaValArgLeu... 293
1074 TTTGAATGAACGTATAAGAACACAGTTTACCGCGGAGGGGTGTTAATC 1123
294 .....AlaGlyAsn... 296
1124 AGTACCGTCCAAAGTTTAAACAACGAGTGAACCTTTCTTTATGATTAAC 1173
297 .....GluArgGlyAlaAsnAsnGlyGlnAsnValThrPhe...Glu 309
1174 GGCACAGGCAACATCATCTATCAACAACATCAACAGGCGGCGGG 1223
310 AsnAsnGlyThrLeuValIleuAspGlnAsnIleAsnGlnGlyAlaGlyG 326
1224 TTTGATTTTGAAGTGAATTTAGGTTCCCGTGAACAGCAAA...A 1270
326 yLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsnGlyIleT 343
1271 CGTGCAAGGCGCGGCTTCATATACATGAGTGAAGACAGTACGTTACTGG 1320
343 hrTrPLeuGlyAlaGlyLysAspValAlaAspGlyLysLysValValTrp 359
1321 AAAGTAAACGCGGTGGCAACAGCGCGCTGCAAAATCGGCAAAAGCAC 1370
:::

```

```

360 GlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyLysGlyTh 376
1371 GCTGCACGTTTCAAGCCAAAGGGAAACCAAGGCTCGATCAGCGTGGCG 1420
376 rLeuGluIleAsnGlyThrGlyValAlaAsnGlnGlyIleuLysValGlyA 393
1421 AGGTACAGTCACTTTTGGATCGACGACGACGATAAAGCAAAACAA 1470
393 spGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspLysLysValGln 409
1471 GCGTTTGAATATCGCTTGTCAGCGCAGGCGGTACGGTGAACATGA 1520
410 AlaPheSerGlnValGlyIleValSerLysArgArgThrLeuValLeuAs 426
1521 TGCGGATATACAGTTCAACCCGCAACACTATTTGCGCTTTCGCGCG 1570
426 nserProAspGlnIleAsnProAsnAsnLeuTyrPheGlyPheArgGlyG 443
1571 GACGTTTGAATTTAAAGCGGATTCGCTTTCGTTCCACCGTATTCMAAT 1620
443 IyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHisIleArgAsn 459
1621 ACCGATGAGGCGCGATGATGNCATCATATGACCAACACATGCAC 1670
460 ValAspGluGlyAlaArgIleValAsnHisAsnThrAspAlaSerTh 476
1671 CGTTACCTTTCACAGGAAATGAAGTATTACCAACGAGTGTGAAGATA 1720
476 rIleThrLeuThrGlyLysSerLeuIleThrAlaProGlnAsnLeuSerV 493
1721 TC 1722
493 al 493

```

seq_name: pir2:S61332

seq_documentation block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (str N:Alternate names: Iga1 protease
C:Species: Haemophilus influenzae
A:Variety: HK635
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria meningitidis strains
A:Reference number: S61314; MUID:95302961
A:Accession: S61332
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-462 <LOW>
A:Cross-references: EMBL:82488; NID:g732714; PIDN:CA51871.1; PID:g732715
A:Experimental source: strain HK635
A:Note: The authors did not translate the codon for residue 462
C:gene: iga
C:superfamily: Iga-specific metalloendopeptidase
C:keywords: hydrolase; metalloprotease

alignment_scores:

Quality:	760.00	Length:	510
Ratio:	2.353	Gaps:	25
Percent Similarity:	63.333	Percent Identity:	38.627

alignment_block:

US-09-303-518d-651 x S61332 ..

Align seg 1/1 to: S61332 from: 1 to: 462

```

202 GCGAAGATTTAGGTTTACGCAAAACAAAGGAGTGTGCGCAATC 251
||| :::::::::::::::::::::
1 AlaThrAsnValGluValArgAspLysAsnHisSerLeuGlyAsnAl 17

```

```

252 AATGACAAAAGCC...CCGATGATGATTTTCTGTGTGTCGCGTAAC. 297
    ::::: ::::: ||||| ||||| ||||| ||||| ::::: |||
17 aleuProAsnGlyIleProMetIleAspSerValIleAspValAsnL 34
298 ..GGCGTGGCGGCGATTGTGGCGGATCATATATTTGAGCGGCGACAT 345
    ::::: ||||| ||||| ||||| ||||| ||||| |||||
34 ysArgIleGlyThrLeuValIleAspProGlnThrIleValSerValLysHis 50
346 AAGCGCGGCTATACACGCTTGATTTT.....GGTGC 377
    ||| ||| ||||| |||||
51 AlaHisGlnIleMetAsn...AspPheTyrPheGlyHisTyrAsnGlyHis 66
378 GGAAGCAAGNAATCCCGATCAGACCGCTTTTCTTACCAAAATGTGAAA 427
    ::::: ::::: ||||| ||||| ||||| ||||| ::::: |||
66 sArgAspValSerAspAspGlnAsnLys.....TyrSerValValThrG 81
428 GAAATATATTAAGCT...GACATTCACACCTTACAAAC.....GGC 468
    ::||| ||||| ::||| ::||| |||
81 IAsnAsnValIleAsnProAsnGlnIleAsnThrPheIleValAspLysArgLeuAsp 97
469 GATTANCAATATCCGCGTTTGCTATTAATTTGTACAGATGACAGAACCTGT 518
    ||| ::||| ||||| ||||| ||||| ||||| ||||| ::|||
98 AspTyrAsnMetProArgLeuAsnLysPheValThrGlnValAlaProth 114
519 CGAAATGACAGATGACATGAGGCGCAAT.....ACCTATTCCGATA 539
    ::||| ::||| ::||| ||||| ||||| |||||
114 r.....ThrProThrLeuAlaGlyAspAspLeuGlnThrTyrLysAspL 129
550 AAGAATAATATCCGAGCGGTGCGGCAATCCGCGTCAGCA..... 597
    ||||| ||||| ::||| ||||| ||||| ||||| |||||
129 ysgLulstYrLeuSerPheValArgValGlyAlaGlyArgGlnLeuVal 145
598 .....CACCACTAATTTGGCGTTATGATGATGACAAACACAGC 632
    ||||| ||||| ::||| ::|||
146 TyrGlnLysGlySerHisHis.....ValGlnAspLysGlnHisG 159
633 C.....GATTATCTCTACTCCGCGGCGATGTTAATTTGGCGCA 670
    ||||| ||||| ::||| ::||| ||||| |||||
159 yGlnAspLeuLysAspLeuSerAlaAlaTyrArgTyrAlaIleGlyGlyT 176
671 ATACACAT.....ATG 681
176 hrProTyrLysGlyIleAsnIleAspProSerGlnSerLysGlyLeu 192
682 CAGGTTGGGGAAT.....AATGGCGTANTTAG 710
    ||||| ||||| ::||| |||||
193 IleGlyPheGlyAspSerArgGlnAspHisValIleAsnSerLysThrLe 209
711 TTGAGCGCGCATGTGCGCCATGCCACGACTATGCGCTATGCCGATTG 760
    ||||| ||||| |||||
209 uLeuSerGlnAsp.....ProLeuThrAsnT 218
761 CAGGTGGCGGAGGCGGATGCGCATGTTTATTTATTTATTTATTTATTTA 810
    ::||| ||||| ||||| ||||| ||||| |||||
218 yrcIValLeuGlyAspSerGlySerProLeuPheAlaPheAspLysGln 234
811 AACATTAATGCTGCTCAACGAGCTTTTACAAACCGCTACCTTATTC 860
    ::||| ||||| ||||| ||||| ||||| |||||
235 GlnAsnLysTrpValPheIleGly.....ProTyrLth 245
861 C.....GGCAGGAAACGCTTCCAGCTG...ATAGCA 892
    ::||| ||||| ||||| |||||
245 rTyTrpAlaGlyTyrGlyLysLysSerTrpGlnGlnLysAsnIleTyrL 262
893 AAGATTGCTTACGATGACATTTACAGAGGATACACATACCATACCTGTT 942
    ||||| ||||| ||||| ||||| ||||| |||||
262 ySerGlnPheThrLysAspValIleuAsnLysAspSerAlaGlyLeuLeu 278
943 TTGAACCGCGCAGTACGACATTTTCTTATCATCAACACACACG 992
    ::||| ||||| ||||| ||||| ||||| |||||
279 .....LysGlyAsnThrGlnTyrAsnThrPheSerAsnGlyAsn.. 291

```

```

993 TACGGGTACGGTAACAGAAACCAAGAAAGTTNCCAAATCCAAAGCTTA 1042
    ||||| ::||| ::||| ::||| ::|||
292 ThrSerMetIleSerAsnGlySerGlnLeuLeuL..... 303
1043 AAGTACAGACAGATCCGACTGTTTGACGAATCTTGAATGAAGATGATAA 1092
    ||||| ||||| ||||| ||||| ||||| |||||
304 .....ValAsnLeuPheAspAsnSer...LysHisThrAsnArg 315
1093 GAACAGATTTCAGCGCGGAGGGGCTTATATCATGATCCGATCCAGTTTAA 1142
    ||| ||| ::|||
316 Gln.....LysAlaAs 319
1143 CAACGGTGAACCTTCTTATGATTTACGGCAACGGCAACCTATCT 1192
    ||||| ||||| ||||| ||||| |||||
319 nTyrGlyLysSerValThrPhe.....GlnGlyAsnGlyThrLeuThrL 334
1193 TATCAAAACATCATCAACAGGCGGCGGCTTTGATTTTGAAGATGAT 1242
    ||||| ||||| ||||| ||||| ||||| |||||
334 euLysAsnSerIleAsnGlnGlyAlaGlyGlyLeuPhePheGlnLysn 350
1243 TTTCAGGTC...TCGCGTGAACAAACGAAACGTGGCAAGCGCGGCGT 1289
    ::||| ||||| ::||| ::||| ||||| ||||| |||||
351 TyrThrValGlnGlySerSerAspAsnIleValTrpAsnGlyAlaGlyI 367
1290 TCATATCAGTGAAGACATGACCTTACTTGAAGTAAACGCGCTGGCAA 1339
    ::||| ||||| ::||| ||||| ||||| |||||
367 eSerValAlaGlnGlyLysThrValThrTrpLysValHisAsnProGlnS 384
1340 ACGACCGCTGTCCAAATCGCAAGGCGACGCTGACCTTCACACCAA 1389
    ::||| ||||| ::||| ||||| ||||| |||||
384 eArgAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlnLys 400
1390 GGGGAAACCAACAGCGCTCGATCGACGCGGCGGACGACGATCTTTGGA 1439
    ||||| ||||| ::||| ||||| ||||| ||||| |||||
401 GlyLysAsnLysGlySerLeuLysValGlyAspGlyThrValIleLeuL 417
1440 TCACAGCGCAGATTAAGGCAAAACCAACCTTACTGTAATCGCT 1489
    ||||| ||||| ::||| ||||| ||||| |||||
417 sGlnGlnAlaAspAlaAsnAsnLysValLysAlaPheSerGlnValGlyI 434
1490 TGNWTCAGCGGCGGCTACGGTGCAACTGAATGCCATATATGATTCAC 1539
    ::||| ||||| ||||| ||||| ||||| |||||
434 LeuValSerGlyArgSerThrValIleuAsnAspAspLysGlnValAsp 450
1540 CCGGACAAACCTCATTTGCGCTTTCGCGGC 1569
    ||||| ||||| ||||| ||||| |||||
451 ProAsnSerIleTyrPheGlyPheArgGly 460
seq_name: p1r2:S61333
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (st
N:Alternate names: Iga1 Protease
C:Species: Haemophilus influenzae
A:Variety: HK284
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61333
R:Domolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61333
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-471 <LOW>
A:Cross-references: EMBL:X82487; NID:q773205; PIDN:CAAS7870.1; PID:q773206
A:Experimental source: strain HK284
C:Genetics:
A:Gene: iga
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

```

alignment_scores: length: 511
 quality: 687.00
 ratio: 2.134
 Percent Similarity: 63.014 Percent Identity: 35.616

alignment_block:
 US-09-303-518D-651 x S61333 ..

Align seg 1/1 to: S61333 from: 1 to: 471

```

202 GCGCAAGATATGAGTNTACACAAAAGGAGGTGGTGGCAATC 251
    ||| .....||| .....||| .....||| .....|||
    1 AlathrsnvalgluValArgaspIysAsnAsnInserLeuGlySer1 17
252 AATGACAAAAGCC...CCGATGATGATTTTCTGTGGTCCGGTAC 297
    .....||| .....||| .....||| .....||| .....|||
    17 aleuProAsnGlyLeuProMetLeuAspPheSerValValAspValAsnL 34
298 . GCGGTGGCGCATTTGGTGGCGATCATATATTTGAGCTGGACAT 345
    .....||| .....||| .....||| .....||| .....|||
    34 ysArGlyLeuGlyThrLeuValAspProGlnThrLeuValSerValLysHs 50
346 AAC..... GCGGCTATACACCTTGA 368
    .....||| .....||| .....||| .....||| .....|||
    51 AlathrsnvalgluValArgaspIysAsnAsnInserLeuGlySer1 64
369 TTTTGGTGGCAAGAAATCCGATCAGACCGTTTCTTACCAAA 418
    ||| .....||| .....||| .....||| .....|||
    65 . . . . .GlyHisArgAspValSerAspArgLysAsnLys.....TyrSerV 78
419 TTGTGAAGAATATATTAACCTGACATTCACACCTTACAA... 465
    .....||| .....||| .....||| .....||| .....|||
    78 alValThrGlnAsnAsnValLysProAsnGlnAspThrPhsValAspLys 94
466 . . . . .GCGATTCATATGCGCGTTTGCATATTTGTCACAGATG 509
    ||| .....||| .....||| .....||| .....|||
    95 ArgLeuAspAspArgLysAsnMetProArgLeuAsnLysPheValThrGluVal 111
510 AGAAGCTGTCGAATGACAGATGACATGAGGCGAAT.....ACCT 550
    : .....||| .....||| .....||| .....||| .....|||
    111 lAlaProThr.....ThrProThrLeuAlaGlyAspAspLeuGluThrT 126
551 ATTCCGATTAAGAAAATATCCGAGCGTGGCCGATCGGCTCAGACAC 600
    ||| .....||| .....||| .....||| .....|||
    126 yrlLysAspLysGlyLysThrProSerPheValArgValGlyAlaGlyThr 142
601 CACTAT.....TGCGGTTA 614
    : .....||| .....||| .....||| .....||| .....|||
    143 GlnPheValThrGluLysGlySerThrValGluLysThrThrArgAs 159
615 TGATGATGACAAACAGCGGATTTATCCACTCCGCGCATGTTAAATTG 664
    : .....||| .....||| .....||| .....||| .....|||
    159 mAsnAspLysPheLeuAspGlnLysLys.....ArgTyrAlaIleG 174
665 GCGGCAATACACATATGACAGGTTGGGAAATATGCGTATTAAGTTTG 714
    ||| .....||| .....||| .....||| .....|||
    174 lGlyLys.....ThrProThrGluGlyIleAsnLysAspProSerLys 189
715 AGCGGC.....GATGT 725
    : .....||| .....||| .....||| .....||| .....|||
    190 LysGlyLeuIleGlyPheGlyAspSerArgLysAsnHisValIleAspAl 206
726 GCGCCATGCGCAAGCACTATGCGCTATGCGATTCAGAGTGGCGAGCG 775
    : .....||| .....||| .....||| .....||| .....|||
    206 alYsThrLeuLeuSerGlnAspProLeuThrAsnThrGlyValLeuGlyA 223
776 ACAGCGGTTCCCAATGTTTATTAAGCAAAACAAACAAATAAATGCGTG 825
    ||| .....||| .....||| .....||| .....|||
    223 spSerGlySerProLeuPheAlaPheAspLysGlnGlnAsnLysThrPal 239
826 CTCACAGGCGGTTTACAAACCGGCTACCTTATTC..... 861
    : .....||| .....||| .....||| .....||| .....|||
    240 PheIleGly.....ProThrThrThrThrThrPalGlyLys 250
  
```

```

862 . GCGAGGAAAMC.....GATTCCAGCTGATACGCAAAAGATTGTTCT 904
    ||| .....||| .....||| .....||| .....|||
    250 rGlyLysSerThrGlnGluThrAsnLysThrLysLysAsp.....PheA 266
905 AGCATGACATTTACAGAGCGGATACACATACCGTCTNTTTGACCGCGC 954
    : .....||| .....||| .....||| .....||| .....|||
    266 lAspAsnLysLysArgAspAsnAlaGluAlaVal.....ProPhe 280
955 AGTAACGACATTTCTTTCATATCCAAACAAACGAGTACGATACGT 1004
    ||| .....||| .....||| .....||| .....|||
    281 SerThrSerGlnThrHisThrPhrAsnThrPhrAsnHisGlnSerGluI 297
1005 AACAGAAACCAAGAAAGGATTCGATTCGAAAGCTTAAAGTACAGAG 1054
    : .....||| .....||| .....||| .....||| .....|||
    297 euLysAsnThrAspHisThrLle.....ThrV 306
1055 TCCGACTGTTTACGATCTTGAATGAACATGAATAAGAACAGTTTAC 1104
    ||| .....||| .....||| .....||| .....|||
    306 alThrLeu.....ProSerAspProAsnArgLeu... 315
1105 GCGGAGGCGGCTTAAAT...CAGTACCGTCCAAAGTTAAACACGCTGA 1151
    ||| .....||| .....||| .....||| .....|||
    316 . . . . .ValAsnPheGlnGlnLysGlnHisLeuGlnThrGlyG 328
1152 AACCTTTCTTTATTCAT.....TACGCAACGCGCAACTCATCT 1192
    : .....||| .....||| .....||| .....||| .....|||
    328 nasValThrPheAspAspSerThrAsnAsnGlyLysGlyThrLeuIleL 345
1193 TATCAAAACATCAACCAAGCGCGGTTGTATTTGAAGTGAT 1242
    : .....||| .....||| .....||| .....||| .....|||
    345 euAspAspHisThrLeuAsnGlnGlyAlaGlyLysLeuPhePheLysGlyAsn 361
1243 TTTACGCTCTGCGCTGAACACAGAA...ACGTGGCAAGCGCGGCGCT 1289
    : .....||| .....||| .....||| .....||| .....|||
    362 TyrGluValLysGlyLysThrAspAspLleThrThrPalGlyGlyLysI 378
1290 TCATTCAGTGAAGACAGTACCGTACTTGAAGTAAACGCGCTGCCAA 1339
    : .....||| .....||| .....||| .....||| .....|||
    378 eaSpValAlaGluGlyLysGluValValThrPlyValHisAsnProGluL 395
1340 ACGACCGGCTTCCAAATCGGCAAGGACGCGTCAAGCTTCAGGCCAA 1389
    : .....||| .....||| .....||| .....||| .....|||
    395 ysAspHisLeuAlaLysIleGlyLysGlyThrLeuIleValGlnGlyLys 411
1390 GGGGAAACCAAGGCTCGATACGCGTGGCGAGCGTACATTTTGA 1439
    ||| .....||| .....||| .....||| .....|||
    412 GlnAsnAsnLysGlySerLeuLysValGlyAspGlyThrValValLeuLys 428
1440 TCACAGGCAACGATTAAGGCAAAACAAACGCTTATGAAATCGGCT 1489
    : .....||| .....||| .....||| .....||| .....|||
    428 sGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSerValGlyI 444
1490 TGATTCAGCGGAGGCTGATGCGTGCACACTGAATGCCATTAATCAGTTTAC 1539
    : .....||| .....||| .....||| .....||| .....|||
    444 lValSerGlyArgSerThrValValLeuAsnAspAspLysGlnValAsp 460
1540 CCCGCAACACTTATTTGCGCTTTCGCGGCGGA 1572
    ||| .....||| .....||| .....||| .....|||
    461 ProAsnSerIleThrPheGlyPheArgGlyGly 471
seq_name: p1r2:T00317
seq_documentation_block:
probable serine protease esp, extracellular - Escherichia coli plasmid p0157
N.alternate names: putative exoprotein precursor
C.Species: Escherichia coli
C.Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Dec-2000
C.Accession: T00317; T42120
R.Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubo
S.; Shingawa, H.
DNA Res. 5, 1-9, 1998
A.Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemo
  
```



```

922 ..... 922
2996 TCATCAATTGACGTAGTGAAGCAAGCAACAAACCGCTGCCAA 3045
      ::::: |||||
923 ..... LysLysHisLeuSerGly 928
3046 AACCTTAATTTCACCTTCGCAAAAGCAACAGTCGATGCCGGCGGTGGCG 3095
      ::::: |||||
929 GLuSnsnLleuLeu..... ValAspPhe..... 937
3096 TTACCAACTATCCGCAAAAGCGCGAGTCCGCTTCATTAATCCGCTCA 3145
      ::::: |||||
938 ..... LeuGlnLysProThr 943
3146 AAGAACAAGACTTTCGACAACTCGCAAGCAGAGCAAAACAAACAG 3195
      ::::: |||||
943 roGluLysGlnLeuAsnLleGlu..... 950
3196 GCGGAAAAAGCAAGCGCAAAAGCTTGACGCGCTGATTCGCGCGCGGCG 3245
      ::::: |||||
951 ..... LeuValSerAlaProLys 956
3246 CGATGCCCGCCAAAGACAGAAACGCTTCCGCAACCGCGCGCGCGCGAG 3295
      ::::: |||||
956 SASPTNrsnGlnSnsnValLpPheLys..... AlaS 966
3296 GCGGGGAAATGTGCGCATTAATGACGCGGAGAGAAAGAAAAAGGGGTG 3345
      ::::: |||||
966 eLysGlnThrLleGlyPhe..... 972
3346 CAGCGGATAAAGACAGCGCNTTGGCGAAACGCGGCAACCGGAAACCG 3395
      ::::: |||||
973 ..... SerAspValThr 976
3396 GCCGNTACACCGCGCTTCCCCCGCGCGCGCGCGCGGGATTTGC 3445
      ::::: |||||
976 PProValLleThr..... 981
3446 CGCAACCGCAGCCCAACCGCACTCAACCCCAACCGCAGCGGACCTG 3495
      ::::: |||||
982 ..... ArgGlnThr 984
3496 ATNACCGCTTATGCCAATAGCGGTTGAGTAATTTCCGCGAGCGCTAA 3545
      ::::: |||||
985 AspAspLysLleThrTrpSer... LeuThrGlyTrpAsnThrValAlaAs 1000
3546 CAGCGTTTTCGCGCTACAGAGCAATTTGACGCGCTTTGCGCGCAAGAC 3595
      ::::: |||||
1000 nLys..... GlnAlaThr 1004
3596 GCCGAAACGCGNTTGGCAACAGCAGCAGCGGNAACCAACACTACCGT 3645
      ::::: |||||
1004 hrArgSnsnLalaLalaLeuPheSerVal..... 1013
3646 TCGCAAGATTTCCGCGCTAC..... CGCA 3671
      ::::: |||||
1014 ..... AspTyrLysAlaPheLeuAsnGlnValAsnAsnLeuAsnLysAr 1028
3672 ACAACCGCAGTTCGCAAACTGATGCAAAAAACCTC..... 3711
      ::::: |||||
1028 gMetGlyAspLeuArgAspLleAsnGlnLysLalaGlyAlaThrPalAlaArg 1045
3712 ..... GCGAGCGCGCGCGTTCGCGCATCTGTTTCGCAACCGGAGC 3753
      ::::: |||||
1045 leMetSerGlyThrGlySerAlaSerGlyGlyPheSerAspAsnTyrThr 1061
3754 GAAACANCTTGACGACGCGCATCGCAACTCGGACGCGCTTGCACAGC 3803
      ::::: |||||
1062 His..... ValGlnValGlyValAspLysLysHisGlnLeuAspGlyLe 1076
3804 CGCGCTTTTCGCGCATACGCGCATCGGAGGTTGCGCATCGGCAGC 3849
      ::::: |||||
1076 unspleuphnerhGlyPheThrValThrHisThrAspSerSerAlaSerA 1093

```

```

3850 ..... AGCAGCGCGCGGGT..... 3864
1093 laAspValPheSerGlyLysThrLysSerValGlyValGlyLeuThrAla 1109
3865 ..... TTACGACGCGCANTCTNTCAGACGCGCATCGAGGCAAAAT 3905
      ::::: |||||
1110 SerAlaMetPheAspSerGlyAlaThrLysLysLeuLleGly..... 1123
3906 CCGCGCGCGCGGTGTCATTTACGCGCATTCAGGACAGATACCGCGCGGT 3955
      ::::: |||||
1124 ..... LysTyrValHisHis..... AspAsnGlnTyrThrAlaThrP 1136
3956 TCGCGCGATTCGCGCATCGAACCGTCAATCGGCGCAACGCGCATTTGTC 4005
      ::::: |||||
1136 heLalaGlyLeuGlyThrArgAspTyr..... 1144
4006 CAAAAAGCGATTCACGCTACGAAAAACGTCAATATGCGCACCCCGGTCT 4055
      ::::: |||||
1144 ..... 1144
4056 TCGCTTCAACCGNTACCGCGCGCATTAAGCAGATTAATTCATTCAAAC 4105
      ::::: |||||
1145 .SerThrHisSerTrpTyrAlaGlyValGlnLalaGlyTyrArgTyrHisV 1161
4106 CCGCGCAACACATNTCCATCAACNCTTATTNAGCCTGCTCATACCGAT 4155
      ::::: |||||
1161 alThrGlnAspAlaThrPileGluProGlnAlaGlnLeuValThr... Gly 1176
4156 GCCGCTTCGCGCAACGTC..... 4173
1177 SerValSerGlyLysGlnThrAlaThrPlyAspGlnGlyMetHisLeuS 1193
4174 ..... CGAACACCGCTCAATA 4189
1193 rMetLysAspLysAspTyrAsnProLeuLleGlyArgThrGlyVal... 1208
4190 CCGCGNTATTTGCTCAGGATTTGCGCAAAACCGCAGTGC... GAATGG 4236
      ::::: |||||
1209 ..... AspValGlyLysSerPheSerGlyLysAspTrp 1219
4237 GCGTAAACGCC 4248
      ::::: |||||
1220 LysValThrAla 1223
seq_name: p1r2:B41500
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 2 - Haemophilus influenzae (fra
N:Alternate names: immunoglobulin A1 proteinase type 2
C:Species: Haemophilus influenzae
C:date: 30-Jun-1992 #sequence-revision 30-Jun-1992 #text-change 08-Dec-2000
C:Accession: B41500
R:Grundy, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A:title: Localization of the cleavage site specificity determinant of Haemophilus inf
A:reference number: A41500; MUID:90129281
A:Accession: B41500
A:status: not compared with conceptual translation
A:molecule type: DNA
A:Residues: 1-384 <GRU>
C:comment: This Iga1 proteinase is classified as type 2 because it cleaves at a proli
C:superfamily: Iga-specific metalloendopeptidase
C:keywords: hydrolase, metalloproteinase
alignment_scores:
Quality: 444.50 Length: 426
Ratio: 1.837 Gaps: 22
Percent Similarity: 56.808 Percent Identity: 32.160
alignment_block:
us-09-303-518d-651 x B41500 ..

```

Align seg 1/1 to: B41500 from: 1 to: 384

```

64 GGGTTCGGCTGCTACTTACCATATGCGTGTGGCATTCCTGCC 113
   |||||
5  LysPheIysLeuAsnPhelIeAlaLeuThrValAlaIatyrAlaLeuThrPr 21
   |||||
114 CCAAGCTTGGCGGACACACTTATTCGGCATCACTACCACTACTATTC 163
   |||||
21 oTyrThrGluAlaAlaLeuValIarGAspAspValAspTyrGlnIlePheA 38
   |||||
164 GCGACTTTCGCGAAATTAAGCAAGTTTGCATCGGGCGGCAAGATATT 213
   |||||
38 rGAspPheAlaGluAsnLysGlyIarGpPheSerValGlyIatThrAsnVal 54
   |||||
214 GAGGTATACACAAAAAGGGGTTGGTCGCGCAATCAATGACAAAGC 263
   |||||
55 GluValIarGAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
   |||||
264 C...CGGATGATGATTTTCTGTGTGTGCGGTAC...GGCGTGGCGG 307
   |||||
71 yIleProMetIleAspPheSerValAlaAspValAsnLysArgIleGlyT 88
   |||||
308 CATTTGGGGCGCATCAATATATGTGAGCGTGCACATACGGCGGCTAT 357
   |||||
88 hrIeuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104
   |||||
358 AACACAGTTGATTT...GGTGGGAGAGAAAGNA 389
   |||||
105 MetAsn...AspPheTyrPheGlyHisTyrAsnGlyHisArgAspValSe 120
   |||||
390 TCCCGATTCAGCACCGCTTTTTCACCAATGTGAAAGAAATTAATTATA 439
   |||||
120 rAspAspGluAsnLys...TyrSerValValThrGlnAsnAsnValA 135
   |||||
440 AGCCT...GACAATTCACACCTTACAAAC...GGCGATTACATATG 480
   |||||
135 snProAsnGlnAsnTrpHisValAspLysArgLeuAspAspTyrAsnMet 151
   |||||
481 CCGCGTTTCATTAATTTGTCTACAGATGACAGACTGTGCAAAATGACAG 530
   |||||
152 ProArgLeuAsnLysPheValThrGluValAlaProThr...ThrPr 166
   |||||
531 TGCATGAGAGGGGAAT...ACCTATTCGATTAAGAAAAATATC 571
   |||||
166 oThrIeuAlaGlyAspAspLeuGlnThrTyrLysAspLysGlnLysGly 183
   |||||
572 CCGAGCGTCCGATCGGCTCAGA...GGCGATTACATATG 597
   |||||
183 euSerPheValIarValGlyAlaGlyArgGlnLeuValTyrGlnLysGly 199
   |||||
598 ...CACCACTATTTGGCTTATGATGATGACAAACACGGC... 633
   |||||
200 SerHisHis...ValGluAspLysGlnHisGlyGluAspLeuLys 213
   |||||
634 GATTATCTACTACCGCGCATGTTAATGGCGGCAATACAT... 678
   |||||
213 sAspLeuSerAlaIatyrArgTyrAlaIleGlyIleThrProTyrLysG 230
   |||||
679 ...ATGCAGGGTGGGGA 693
   |||||
230 yIleAsnIleAspProSerGlnSerLysGlyLeuIleGlyPheGly 246
   |||||
694 AAT...AATGGCTAATTAAGTTTGAAGCGGCA 722
   |||||
247 AspSerArgGluAspHisValIleAsnSerLysThrLeuLeuSerGlns 263
   |||||
723 TGTGCGCATGSCAAGCACTATGGCCATGCGGATTCGAGTGGCGGAG 772
   |||||
263 P...ProLeuThrAsnTyrGlyValLeuG 272
   |||||
773 GCGACAGCGGTGCGCAATGTTATTATGACAAACAAACATTAATG 822
   |||||

```

```

272 LysAspSerGlySerProLeuPheAlaPheAspLysGlnHisnLysTyr 288
   |||||
823 CTGCTCAACGAGATTTTACAAACCGGCTACCTTATTCGCGCAGGAA 872
   |||||
289 PheIleIleGlyProTyrThrTyrTrpAlaGlyTyrGlyLysIleSerH 305
   |||||
873 CGGTTTCACAGCTGATACGCAAGATTTGTCTGATGATTCATTTACAG 922
   |||||
305 rGlnGlnIleTrpAsnIleTyrLysSerGlnPheThrLysAspValLeuAsn 322
   |||||
923 GCGATACACATACCGCTCTNTTTGAACCGGCGAGTACAGGACAT... 966
   |||||
322 yAspSerAlaGlyLeuLeu...LysGlyHisThrGln 333
   |||||
967 TTTTCCCTTATCATCCAAACACAGGTACGGGTACGAGCAACCA 1016
   |||||
334 TyrAsnTrpThrSerAsnGlyAsn...ThrSerMetIleSerAsnGlySe 349
   |||||
1017 CGAAAGGCTTCCATCCAAAGCTTAAGTACAGACAGTCCGATGTTG 1066
   |||||
349 rGluLeuLeuGly...ValAsnLeuPheA 358
   |||||
1067 ACCAATCTTTGATGAACGATTAAGACACAGTTTACGGCGCAGGGGT 1116
   |||||
358 spAsnSer...LysHisThrAsnArgGlu... 366
   |||||
1117 GTTAATCAGTACCGTCCAAAGCTTAACACAGGTGAAACCTTCTTTTAT 1166
   |||||
367 ...GlnGlyAsnGlyThrLeuThrLeu 384
   |||||
1167 CGATTACGCAACGCGCAACATCATCTTA 1194
   |||||
377 ...GlnGlyAsnGlyThrLeuThrLeu 384
   |||||

```

seq.name: p1r2:A65044

seq.documentatn_block:

hypoetical protein b2647 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999

C:Accession: A65044

R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

.A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; M01D:9742617

A:Accession: A65044

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1569 <BLAT>

A:Cross-references: GB:AE000350; GB:U00096; NID:92367147; PIDN:AMC75695.1; PID:g17894

A:Experimental source: strain K-12, substrain Mgi655

alignment_scores:

Quality:	372.50	Length:	1676
Ratio:	0.491	Gaps:	76
Percent Similarity:	45.286	Percent Identity:	19.153

alignment_block:

US-09-303-518D-651 x A65044 ..

Align seg 1/1 to: A65044 from: 1 to: 1569

```

127 GCACACACTTATTTGGCATCACTACCAATACTATCGGACTTTCGCA 176
   |||||
160 GlyThrThrIleGluSerGlyAsnGlnAspValTyrLysGlyIleSe 176
   |||||
177 AAATAAAGGCAAGTTTGCAGTCCGGCGGCAAGATATGAGTNTACACA 226
   |||||
176 rAsnGlyThrThrIleLysGlyIleAsnArgValGlnGlyGlySerA 193
   |||||
227 AAAAAAGGAGTTGGTC...GGCAATCAATGACAAACCCCG... 267
   |||||

```



```

1828 GAGAGCCGACCCNGCTGCTTCCGGGAGACAATTTAAACGGCAACAT 1877
      ::::::::::: :::: ::::: ::::: |||
672 SerAspLysAlaAsnIleLysGlyLysGlnThrValTyrGlyLeuAl 688
1878 CAGCAACAACAAC.....GCCAAACGTGTTTTCAGCGGACGACCGA 1918
      ||::::::::||| ::::: ::::: |||
688 aThrGluAlaAsnIleGlnSerIleGlnIleValAspGlyGlySerI 705
1919 CACCGCAGCGCTTACATTCATTAGAACGGCGTGTCAAAAATGGAAGGT 1968
      || ::::::::::: ||| ::::: :::::
705 hrgLysLys.....ThrHisIleAsnGlyLysThrGlnThrValGlnAsn 719
1969 ATCCCAACAAGA.....GAATCGTGTGGACACAGACTGATCA 2009
      ::::: ::::: ||| :::::
720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnGlnIle 736
2010 CCGCAGCTTTAAAGCGGAAAATTCATATTCAGCGCGGACGCGGTGA 2059
      ::::: ||::::::::: ||::::::::: |||
736 tAlaAsnGlyThrAlaGlnGlySerIleIleAsnGlyLysGlnVal 753
2060 TTTCCCGCG..... 2067
      :::::::::::
753 aLAsnGlnGlyLysLeuAlaGlnAsnSerValLeuAsnAspGlyLys 769
2068 ...AATGTGCCAAAGTGAAGCGGATTGNCATTTGAGCAATCAGCCCA 2114
      ::::: ::::: ||| :::::
770 LeuAspValAlaArgLysGlySerAlaThrGlyIleGlnGlnSerSer 786
2115 ACAGATTTTGGTGCACGCGATCAACGCAATCACTATACACTT 2164
      ||::::::::: ::::: ::::: |||
786 nGlyAlaLeuValAlaIleThrThrArgAlaThrGlyThrArgA 803
2165 CGAGCTGACN.....GCTGTGACAAAT..... 2187
      ::::: ::::: ||| :::::
803 lAspGlyAlaIlePheSerIleGlnGlnGlyAlaAlaAsnIleLeu 819
2188 .....TGTCGCAANAANCAATTACCGACGATPA 2216
      ||| ::::: |||
820 LeuAlaAsnGlyLysValLeuThrValGlnSerAspThrSerSerAsp 836
2217 AGTGATGCTTCATG.....ACTAAGANGACN 2245
      ||::: ::::: ::::: |||
836 sThrGlnValAsnMetGlyLysArgGlnIleValLysThrLysAlaThr 853
2246 TTAAGCGCANGTGTAAGCTNNCCATNACGNTNNTTNAANCTCNCGG 2295
      ::::: ||| :::::
853 lAthrGly.....ThrThrLeuThrGly 861
2296 CNTGCNCACTNAANGC..... 2313
      ::::: |||
862 GluGlnIleValGlnGlyValAlaAsnGlnThrThrIleAsnAspGly 878
2314 .....AATCTTAGTGCAAATGGCGAT.....ACAGTTATACAG 2347
      ::::: ||::::::::: |||
878 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsn 895
2348 TCAGCCACAAGCCACCCAAAACGCAAC.....CTTACGCTGTG 2388
      ::::: ||| ||| :::::
895 lGlnGlyLysThrLeuThrValAsnAspAsnGlyLysAlaThrAspIle 911
2389 GGCATATGCCAAGCAACATTAATCAAGCCATTAACGCGCAACNCATC 2438
      ||::: ||::::::::: |||
912 GlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly...Ile 927
2439 GGNTTGGGCAAT.....GCTTCAATTATCTAAGCAACAACGCG 2479
      ::::: ||| :::::
927 uIleSerGlyThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeu 944
2480 CACAACAAC.....GGCAGTGTGAGCGTTTCGACAAAC 2511
      ||| ||| :::::
944 lAthrAsnMetLeuLeuGlnLysAsnGlyLysAsnLeuValLeuAlaGly 960

```

```

2512 GCTAAGCGCAACGTAAGCATTCGCCACTCAACGCAATGTCTCCTAGC 2561
      ::::::::::: ||| :::::
961 ThrGluAlaArgAspSerThrValGlyLysGly..... 972
2562 CGATAGGCGATATTCATTTTGAANAACGCGGCTTACCGGACACTCA 2611
      ||::: ::::: ::::: |||
973 .....AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsn 987
2612 GCGGACGACAGGAAACACCATTAACCTTAAAGACGCAATGACGCTG 2661
      ||::: ||| :::::
987 eGlyGlyGlnTyrThrLeuGlnArgSerLysAspGluPheGlnAlaLeu 1003
2662 CCGTAGCGACGGAATTA..... 2679
      ::::: |||
1004 AlaArgAlaGlnAspLeuGlnValAlaGlyLysThrAlaIleValTyrAl 1020
2680 .....GCCAATTAAAC 2692
      :::::::::::
1020 aGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSer 1037
2693 TT..... 2694
      ||
1037 eumEThrProArgAspAsnValThrProValLysLeuGlnGlyAlaVal 1053
2695 .....GACAAGCGCACATTAACCTCAATCCGCTAT..... 2727
      ||::::::::: ||::::::::: |||
1054 ArgIleThrAspSerAlaThrLeuThrLeuGlnLysAsnGlyAlaSprThr 1070
2728 ..CGCAGACATCTGCACGCGGCAACCGCAGAGTG..... 2763
      ||| ::::: |||
1070 rLeuAlaSprLeuThrAlaIleSerArgGlySerValThrLeuAsnSer 1087
2764 .....TCAGACAGCGCGCGCGCTTCCG.....CGTTCCTA 2799
      ::::: ||| :::::
1087 snAsnSerCysAlaGlyThrSerAsnGlyLysThrValAsnSerLeu 1103
2800 TTA.....TCGCTTACCGCGCAAC 2819
      |||
1104 LeuLeuAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaAlaProAl 1120
2820 TTCCGTTGAATCCCGTTTCAACAGCTGACGTAACGCAAAATTGAACN 2869
      ::::: ||::::::::: |||
1120 aThrThrAsnGlyLysThrAsnThrLeuThrAsn...GlnLeuSerG 1136
2870 GTCAGGACATCCGCTTATGTGCGAAGCTTTCGCTACCGCAAGCAG 2919
      ||::::::::: |||
1136 LysGlyAsnPheTyrLeuHisThrAsnValAlaGlySerArgGlyAsp 1152
2920 AATTTGAAGCTGGCGGAAGCTCCGAAGNACTTACACCTTGCGGCTCA 2969
      ::::: ||| :::::
1153 GlnLeuValValAsnAsnAlaThrGlnLysAsnPheLysIlePheValG 1169
2970 CAATACCGGCAAGAACCCGTAAGCTCGATCAATTGACGGTAGTGAA 3019
      ::::: ||| :::::
1169 nAspThrGlyValSerProGlnSerAspAspAlaMetThrLeuVal... 1184
3020 GGAAGACACAACACCGCTGTCCGAACCTTAATTCACCCGCAAAAC 3069
      |||
1185 .....LysThrGlyGlyLysAlaSerPheThrLeuGlyAsn 1197
3070 GAA.....CAGCTGATGCGGCGGCGGTGAGCTTACCACTACGCGAA 3113
      ||| |||
1198 ThrGlyLysPheValAspLeuGlyThrTyrGlnTyr...ValLeuLys 1213
3114 AGACGCG.....GAGTCCGGCTGATATCCGCTCAAGAAGAAGAC 3157
      ::::: |||
1213 lAspGlyAsnSerAsnThrPheAsnLeuThrAsnAspValLys..... 1226
3158 TTTCGACAACACTGGCAAGGACAGAAACCAAAAACAGCGGAAAAAG 3207
      :::::
1226 ..... 1226
3208 AACGCGCAACGCTTGACCGCGCTGATGGCGCGGCGGATGCCGCGA 3257

```

```

1227 .....Pr 1227
3258 AAGACAGAAAGCGTTGCCAACCGCGCGCGGAGGCGGGAATG 3307
1227 oAsnProAspProIleProAsnPro..... 1235
3308 TCGGCATTATGACGGCGGAGGAAAGAAAAACGGGTGCAGCGGATAA 3357
1235 ..... 1235
3358 GACAGCCNTTGGGGAACAGCGGAAACCGGCGGNTACAC 3407
1236 .....LysProAspProIysProAspProIy 1244
3408 CGCCTCCCGCGCGCGCGCGCGCGCGGATTCGCCAACCGCAGC 3457
1244 sProAspProAsnProIys.....ProAspProIthP 1255
3458 CCCAACGCCAACCTTCACCCCAACCGGCGGACCTGATNAGCGTTAT 3507
1255 roAspProIthPProIthProAlProIulysArgIlethPro..... 1269
3508 GCCAATAGCGGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTC.. 3555
1270 SerThrAlaValLeuAsnMetAlaIaThrLeuProLeuValPheAs 1286
3556 .....GCCGACAGCAGCAATGAGCCGCTGTTCCGACAG 3592
1286 palagIuLeuAsnSerIleArgIuArgLeuAsnIleMetLysAlaSerP 1303
3593 ACCGCGGACGCGNTTGGACAGCNGCATCCGACACCAACACATAC 3642
1303 roHIsAsnAsnValIthProIylalThrIyAsnThraAsnAsnVal 1319
3643 CGTTGCGAAGATTTCCGCGCTACCGCCACAAACGACCTGCGCAAT 3692
1320 ThrThrAspAlaGlyAlaGlyPheGluGlnThrLeuThrGlyMetIhVa 1336
3693 CGGATGACAGAAAAAC.....CTCGGAGCGGCGCGCTGCGCA 3730
1336 IgIyIleAspSerArgAsnAspIleProGluGlyIleThrThrIleuGlyA 1353
3731 TCCTGTTTCCACACACGAGCAGCAACANTCTGACGAGCGATCGGC 3780
1353 lArPheMeGlyTyIserIleSerHisIleGlyPheAspArgIy..... 1367
3781 AACTCGACAGCGCTTGGCCACGCGCGCTTTCGGGCATACGCGATCG 3830
1368 .....GlyHisGlySerVal...GlySerTyIserLeuG 1378
3831 C..... 3831
1378 yGlyTyIAlaSerTrpGluHisGlySerGlyPheTyIleuAspGlyVal 1395
3832 .....AGTTGACATCGGCATC.....AGCAGCGGC 3858
1395 alIyLeuAsnArpPheLysSerAsnValAlaGlyLysMetSerSerGly 1411
3858 GCGGTTTACGACGCGCANTCTNTCAGACGATCGAGCAAAATCG 3908
1412 GlyAlaAlaAsnGlySerTyIHisSerAsnGlyLeuGlyGlyHis.... 1426
3909 CGCGCGCGTGTGCATTAAGGCATCAGCAGATACCGCGCGGTTTC 3958
1427 .....IleGluThrGlyMetArgPheThrArgP 1436
3959 GCGGATTCGCGATCGAACGCTATCGCGCAGCGCTATTTCGTCAA 4008
1436 lYAsnTrpAsnLeuThrProTyIAlaSerLeuThrGlyPheThrAlaAsp 1452
4009 AAAGCGGATTACCGCTACGAAAC.....GTCAATATCGCCAC 4046
::: ::::: ::::: :::::

```

```

1453 AsnProGluTyIHisLeuSerAsnGlyMetLysSerIysSerValAspTh 1469
4047 CCCGCTTCTGCTTCACACGNTACCGN...GCGGCGATTAAAGCAGATT 4093
1469 lArgSerIle.....TyIArgGluLeuGlyAlaIaThrLeuSerT 1482
4094 ATTCATTCAAACCGCGCACACATNTCCATCACNCTTATTNAGCCTG 4143
1482 yAsnMetArgLeuGlyAsnGlyMetGluValGluProTyIleuLysAla 1498
4144 TCCTATACCGATGCCGCTTCGGCAAGTCCGACACCGCTCATACCGC 4193
1499 AlaValArgGlyGluPheValAspAsnAsnArgValLysValAsnSerAs 1515
4194 NGTATTGGCTGAGATTTC.....GGCAAAACCGCGCATGCGAATGG 4237
1515 pGlyAsnPheValAsnTyIleuSerGlyArgArgGlyIleTyIArgIaG 1532
4238 GCGTAACGCGCAAAATCAAAGTTTCACGCTGCTCCNTCCAGCGTCCGCC 4287
1532 lYleLysAlaSerPheSerSer...ThrLeuSerGlyHisLeuGlyVal 1547
4288 GCCAAAGNCCGCAACTGGAGACGCACACAGCGCGCATCAAA..... 4332
1548 Gly.....TyIserHisSerAlaGlyValGluSerPr 1558
4333 .....TTAGCTACCGCTCG 4347
1558 oTrpAsnAlaValAlaGlyValAsnTrp 1567

```

seq_name: p1r2:C91068

```

seq_documentation_block:
  hypothetical protein ECs3515 [imported] - Escherichia coli (strain O157:H7, substrain
  C:Species: Escherichia coli
  C:date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
  C:Accession: C91068
  R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
  gasawara, N.; Yasunaga, T.; Kuahara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
  DNA Res. 8, 11-22, 2001
  A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
  A:Reference number: A9629; PMID:21156231; PMID:11258796
  A:Accession: C91068
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-1571 <NAV>
  A:Cross-references: GB:BA000007; PIDN:BA836938.1; PID:913362986; GSPDB:GN00154
  A:Experimental source: strain O157:H7, substrain R1MD 0509952
  C:Genetics:
  A:Gene: ECs3515

```

```

alignment_scores:
  Quality: 355.00      Length: 1667
  Ratio: 0.462        Gaps: 79
  Percent Similarity: 46.131  Percent Identity: 19.376

```

alignment_block:

us-09-303-518d-651 x C91068 ..

Align seg 1/1 to: C91068 from: 1 to: 1571

```

148 AACTACCAATACATATCGCGACTTTCGCGAAATAAAGGCAAGTTGCAGT 197
111 ::::: :::::
167 AsnGlnAspValTyIArgGlyGlyIleSerAsnGlyThrThrIleLysG 183
198 CGGGCGCAAGATATTGAGGTNTACAAACAAAAGGAGAGTTGTC.... 243
183 yGlyAlaSerArgValGluGlySerAlaAsnGlyThrLeuIleAspG 200
244 ..GGCAATCAATGACAAAAGCCCG...ATGATGATTTTCTGTGGTG 288
111 ::::: :::::
200 lYgIySerGlnIleValIyGlnGlyIleHisAlaAspGlyThrThrIle 216

```

```

289 TCCTGTAACGGCGGTGGCGCATTTGTTG..... 315
217 Asnlyserglyserglinspvalvalglnglyserleualathrsnth 233
316 .....GGCGATCAATATAT.....GTGA 334
233 rthrileasnglylgyrglintyvalgluglnserthrvalglutht 250
335 GCGTGCACATAACGGCGGTATACACAGCT..... 366
250 hrthrileysasnnglygluglnrvalyrgluserargalaleu 266
367 .....GATTTGGTGGCGAAGACNAATCCGATCAGCACCG 404
267 AspThrThrIlegluglythrnglnserleuasnseryserthrhl 283
405 TTTTCTTACCAATTTGTGAAGAATAATTAAGCTGACAAATTCAC 454
283 alysasnthrnglnleryserlgylythrnglnlleleaspsnthr 299
455 ACCCTTACAACGGCGATNACATATGCCGCTTGCAATTTGTCACA 504
300 .....Ser 300
505 GATGCAAACTGTGAAATGACAGT.....GACATGAGGG 542
301 SerSerAspValIlegluValtyrserglylvalleuaspsvalsercl 317
543 GAATACCTTTCGATTAAGAAAATATCCAGCGTCCGATCGGCT 592
317 ygltythralthrsnvalthrln..... 325
593 CAGGACACCATATTGGCGTATGATGATGACAAACGCGATTTATCC 642
325 ..... 325
643 TACTCCGGCGCATGTTAATGGCGCAATACAT.....ATGCA 683
326 Hisaspgllyala...IleleulysThrnsnThrnsnnglyThrThrValse 341
684 GGGTGGGGAATTAATGGCGTANTTAGTTG..... 714
341 rgltyhrnsnsergluglyalapheserIleHisnHisvalalaaapa 358
715 .....AGCGCGATGTGCGCATGCCACAGCACTATGCGCCT 750
358 snvalleuenglunasnglyglHisleuasprlleasnalatygly... 373
751 ATGCCGATTGCAGTGGCGGACGCGCATGCCAGCTTCCCAATGTTATTGA 800
374 .....serlaasnlystrhrillelly 381
801 TGACAAAACAAATTAATGCTGCTCAACGAGCTTTACAAACCGGCT 850
381 sasprysgltyhrmetserValleuthrsnalalysalaspala... 396
851 ACCCTTATCCGGGACGAAACGTTTCCAGCTGATACGCAAGATTGG 900
397 .....ThrarglleaspsnnglylValmet..... 405
901 TTCTACGATGACATTACAGAGCGATACATACCGCTNTTTTGAACC 950
406 .....Aspvalalagllyasnalahrsnthrillelle..... 416
951 GCGCAGTACGACATTTTCTTTTACATCCAAACAGCTACGGGTA 1000
417 .....Asnnglygly...ThrGlnAsnIleasnasnlyrglyleat 430
1001 CGGTAAACAACAACGAAAGTNTCCATCCAAAGCTT.....AAA 1044
430 hrnglythrnsnleasnserglythrlnsnllelysserglyglylys 446

1045 GTACAGACATCCGACTGTGTTGACAAATCTTGAATGAACATGATTAAGA 1094
447 AlaasprThrThrIleleserSerglyserhrnglnvalalglulysas 463
1095 ACCAGT.....TACGGGACAGGGGTGTTAATCAGTACC 1129
463 poltyhrnalaleglyserasnIleaserlalegly..... 475
1130 GTCAAGGTTAAACAAGGTGAAAACCTTCTGTTATGATTCAGGCAAC 1179
476 .....serleullevaltyrthrcly 482
1180 GCGAATCATCTTATCAACACATCAACAGCGCGCGGCTTTGTA 1229
483 Gly.....Ilealanhsglyvalasnnglnuglthrnglyserlale 496
1230 TTTGAAGTGATTTTACGGTCTCGCCCTGAAACACAAACGTCGCAAG 1279
496 uValalasn.....Thrg 501
1280 GCGCGGCGCTTCATATCAGT.....GAAGACATACCGTT 1314
501 lyalaglyThrAsprIlegluglytyrnsnlyserHisphethrIle 517
1315 ACTTGGAAGTAACGGCGGTGCAACAGCGCGCTGTCCAAATGCGCAA 1364
518 Thr.....Glyglyglualalasnlyrvalalvalleuglu...Asnth 530
1365 AGGACGCTGCAGCGTTCAGAACCAAGGGGAAACCAAGGCTGCATCAGCG 1414
530 rgllyleuthrvalalalalysThrserlalyasnthrthrIlea 547
1415 TGGCGCAGGTACAGTATTTGGATCAGCAGCGCAGATTAAGGCAA 1464
547 spalaglyglylysleuIlevalglInlysglnalalys..... 559
1465 AAACAAGCCTTATGTAATCGCTGNTCAGCGCGGAGGTACGGTCA 1514
560 .....ThrAsperthrAr 564
1515 ACTGATATCCGATATATCAGTTCAACCCGACAAACCTATTTGCGCTTC 1564
564 glleuasn.....A 567
1565 GCGGCGACGTTTGATTTAAACGGGATTCGCTTCGTCACCGATT 1614
567 snnglyglvalleuglnvalglInaspelyglyglualalysHis...Val 582
1615 CAAATACCGATGAAAGGCGCATGATGNCNATCATTAATCCCAACAC 1664
583 gluglnlnserglylalelle.....Alaserthrth 595
1665 ATCCACCGTTACATTTACAGGGAATGAACT.....A 1696
595 rSerglyThrleuIlegluglyThrnsnserTyrglyAspalaPhetyrl 612
1697 TTACACAACGAGGTGTAAGATATCATATGATTAATTAACACCAAGAA 1746
612 leargansnserglualalalysasnvalalvalleuglnsnalaglyserleu 628
1747 ATTGCTTACACAGGTTGGTGGCGAAGAAATACGACCAACAAAGACGG 1796
629 ThrValValThrlyserargalalalaspThrIlelleasnalasnol 645
1797 GCGGTCACACCTT..... 1809
645 ylyshweaspsvaltyrglylyshspvalglyThrValleuasnservalag 662
1810 .....GTTTACAGCCGCGCGAAGACGACCGCCNGCTGTT 1848
662 lYthrGlnThrIleTyralaserAlaThrSerAspysalasnIlelys 678
1849 TCCTGGGGAACAATTTAAACGGCAACATCACGCAAAACAAC..... 1890

```

```

679 GlyGlyLysGlnThrValIleThrGlyLeuAlaThrGlnAlaAsnIleGluLeu 655
      ::||| ::::: ||| |||:::||||
1891 .GGCAACCTGTTTTCAGCGCAGACCGACGACGCGCTACAAATCATT 1939
      ||::: ::::: ||| |||
695 rGlyGlnGlnIleValAspGlySerThrGluLys.....ThrHis 710
      ::::: ||| ::::: |||
1940 TAGAAGCGGGGTGTCAAAAATGGAAGTATCCACACAGA..... 1980
      ::::: ||| ::::: |||
710 LeasnGlyGlyThrGlnThrValGlnAsnThrGlyLysAlaIleAsnThr 726
      ::::: ||| ::::: |||
1981 GAATTCCTGGGAGACAGACTGGATCNCAGCGCATTTAAAGCGGAAAA 2030
      ::::: ||| ::::: |||
727 AspIleValSerGlyLeuGlnGlnIleMetAlaAsnGlyThrAlaGlu 743
      ::::: ||| ::::: |||
2031 TTTCATATTCAGGGCGGCGCGGTGATTCCCGC..... 2067
      ::::: ||| ::::: |||
743 YSerIleIleAsnGlySerGlnIleValAsnGluGlyLeuAlaG 760
      ::::: ||| ::::: |||
2068 .....ATGTGGCAAGTGGAA 2085
760 LuasnSerValLeuAsnAspGlyGlyThrLeuAspValArgGluLysGly 776
2086 GGCATTGNCATTGAGCAATCAGCCCAAGCAAGTTTGGTGGCACC 2135
      ::::: ||| ::::: |||
777 SerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuValAlaThr 793
2136 GCATCAAGCCATACATCTGTACAGCTGGAGCTGACN..... 2175
      ::::: ||| ::::: |||
793 rArgAlaThrArgValThrGlyThrArgAlaAspGlyValAlaPheSer 810
2176 .....GTCGTGACAAAT..... 2187
810 LeGlnGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyValLeu 826
2188 TGTGTGGAANAANCATTACCGAGATTAAGTGTGCTTCA..... 2229
      ::::: ||| ::::: |||
827 ThrValGlnSerAspThrSerSerAspLysThrGlnValAlaSerThr 843
2230 .....TTGACTAAGACNGACNTNAGCGGAGNTGTAAGCTNNCAATN 2272
      ::::: ||| ::::: |||
843 rArgGlnIleValLysThrLysAlaThrAlaThrGlyThrLeuThrG 860
2273 ACGNTNNTNAANCTCNCNGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2310
860 LysGlyGlnGlnIleValGlnGlyAlaAlaAsnGlnThrThrIleAsn 876
2311 .....GGCAATCTTAGTGAATGGGATACA...CGTTATACACT 2348
      ::::: ||| ::::: |||
877 GlyGlyIleGlnThrValSerAlaAsnGlyLysAlaIleLysThrThr 893
2349 CAGCCACACGCGCCAC.....CAAAAGCGCAACCTTGGC...C 2383
      ::::: ||| ::::: |||
893 eAsnGlnGlyGlyThrLeuThrValAsnAspAsnGlyLysAlaThrAsp 910
2384 TCGTGGCAATGCCCAACCAATTAATCAAGCCACATTAAGCGGCAAC 2433
      ::::: ||| ::::: |||
910 LeValGlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly 925
2434 NCATGGTTCGGGCAAT.....GCTTCATTTAATCTAAGACAA 2474
      ::::: ||| ::::: |||
926 IleGlnIleSerGlyThrHisGlnThrGlyThrPheSerIleSerGly 942
2475 GCGCGACAAAC.....GGCAGTCTGACGCTTTCG 2506
      ::::: ||| ::::: |||
942 rMetAlaThrAsnMetLeuLeuGlnAsnGlyLysAsnLeuLeuValLeu 959
2507 ACAAGCTAAAGCAACGTAAGCCATTCCGACCTCAACGCAATGCTCC 2556
      ::::: ||| ::::: |||
959 rAgLysThrGlnAlaThrAspSerThrValGlyLysGly..... 972
2557 CTAGCGGATAGGCACTATCTTGTGAAAAACAGCCGCTTACCGGACA 2606
      ::::: ||| ::::: |||

```

```

973 .....AlaMetGlnAsnGlnGlyLysAspSerAlaThrLysVal 985
2607 ACTCAGCGGACGAAGANACAGATTACACTTAACAGACGCAATGCA 2656
      |||::: ||| ::::: |||
985 rAsnSerGlyGlyGlnThrThrLeuGlyArgSerLysAspLysPheGln 1002
2657 CGTCGCCGTCAGGCGGAAATTA..... 2679
      ::::: ||| ::::: |||
1002 rAlaLeuAlaArgAlaGlnLysPheGlnValAlaGlyGlyThrAlaIleVal 1018
2680 .....GCCAATTT 2687
      ::::: ||| ::::: |||
1019 rThrAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySer 1035
2688 AAACCTT..... 2694
      ::::: ||| ::::: |||
1035 rSerLeuMetThrProAlaGlyAspAsnValThrProValLysLeuGlu 1052
2695 .....GACAAAGCCACCAATTAAGTTCAGCTTCCGCTAT... 2727
      ::::: ||| ::::: |||
1052 rAlaLeuArgLysThrAspSerAlaThrLeuThrIleGlyLysAsnGlyVal 1068
2728 .....CGCCAGATGCTGACGCGCGCAACCGCGAGGTG..... 2763
      ::::: ||| ::::: |||
1069 rThrThrLeuAlaAspLeuThrAlaAlaSerArgLysSerValThrPhe 1085
2764 .....TCAGACACCGCGCGCGCGCTTCCGCGC...CGTT 2794
      ::::: ||| ::::: |||
1085 rSerAsnAsnSerCysAlaGlyThrSerAsnGlyLysThrArgValAsn 1102
2795 CCCTATTA.....TCGCTTACACCG 2814
      ::::: ||| ::::: |||
1102 rLeuLeuLeuAsnAspGlyAsnValThrLeuSerAlaGlnThrAlaAla 1118
2815 CCACTTCGCTGGAATCCGCTTCAACAGCTGACGCTTAACGCAAT 2864
      ::::: ||| ::::: |||
1119 rProAlaThrThrAsnGlyLysThrAsnThrLeuThrThrAsn...Glu 1134
2865 GAACNGTCAAGAACATTCGCTTATGTCGGAACCTTTCGCTACCGAA 2914
      ::::: ||| ::::: |||
1134 rSerGlySerGlyAsnPheThrLysThrAsnValAlaGlySerArg 1151
2915 GCGACAAATTAAGCTGCGGGAAGTCCGAGCAAGTTCACCTTACCT 2964
      ::::: ||| ::::: |||
1151 rLysPheGlnLeuValValAsnAsnAlaThrGlyLysPheLysIlePhe 1167
2965 GTCACAAATACGCGCAAGACCGCTGATCAATTAATGACGCTACT 3014
      ::::: ||| ::::: |||
1168 rValGlnAspThrGlyLysSerProGlnSerAspAlaMetThrLeuVal 1184
3015 GGAAGGGAAGACACAAACCGCTGCGAAGAACTTAATTTACCTGC 3064
      ::::: ||| ::::: |||
1184 rLysThrGlyGlyLysPheSerIleGlyLysAlaSerPheSerLeu 1196
3065 AAACCGAA.....CACTGATGCGCGCGCTGATCAACACTATC 3108
      ::::: ||| ::::: |||
1196 rLysThrGlyGlyPheValAspLeuGlyThrGlyLysThr...ValLeu 1211
3109 GCGAAAGACGCG.....GAGTTCGCTGATTAATTCGCTCAAGACA 3152
      ::::: ||| ::::: |||
1212 rLysSerAspLysAsnSerAsnThrPheThrAlaAspAlaLysPro 1228
3153 AGAGCTTTCGACAAACCTGCGCAAGGCAAGCAAGCAAGGCGGAAA 3202
      ::::: ||| ::::: |||
1228 rSerProAspProAsn..... 1232
3203 AAGCAAAAGCGCAAGGCTTACGCGCTGATTCGCGCGCGGATGCG 3252
      ::::: ||| ::::: |||
1232 ..... 1232
3253 GCGGAAAAGACAGAAAGCTTGGCGCAACCGCGCGCGCGGCGGGA 3302
      ::::: ||| ::::: |||
1233 .....ProAsn..... 1234

```

```

3303 AAATGGCGATTATGCAAGCGGAGAGAAAAAAGGGTGCAGCGG 3352
1234 ..... 1234
3353 ATAAAGACAGCGCTTTGGCGAAGCGGCAAGCGGAACCGCGCGGNT 3402
1235 .....ProAsnProAs 1238
3403 ACCACCGCTTCCCGCGCGCGCGCGCGCGGATTTGGCGGACAC 3452
1238 nProLysProAsp..ProLysProAspProLysPro..Asp...ProLysPr 1253
3453 GCAGCCCAACCGCAACCTCACCACCGGACGCGACCTGATNAGCC 3502
1253 oAspProThrProGluProThrProThrProThrProGluLysArgLle 1270
3503 GTTATGCGCATAGCGGTTTGAAGATTTTCCGCGACCTCAACAGCTT 3552
1270 hrProSerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuAl 1286
3553 TTC.....GCCGTACAGAGCATGTGACCGCGCTGTTCG 3587
1287 PheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAl 1303
3588 CGAAGACCGCGCGCAACGCGCTTTGACAAACGCGCATCCGGNACCAAC 3637
1303 aSerProHisAsnAsnValAlrpgLysAlaThrLysAsn...ThrArgA 1319
3638 ACTACCGCTTCCCGCGCTTCCGCGACCGCAACACGCGCGCTGCGC 3687
1319 snAsnValThrThrAspAlaGlyAlaGlyPheIleuGlnThr..... 1332
3688 CAAATCGGTATGCAAAAAACCTC.....CG 3713
1333 LeuThrGlyMetThrValGlyLeaSerProAsnAspLleProGlu 1349
3714 CAGCGGCGCGGTGGCATCTGTTCGCAACACGCGCGCAAAACANT 3763
1349 yIleAlaThrLeuGlyAlaPheMetGlyLysSerHisSerHisIleGly 1366
3764 TCGACGACGCGCATCGGCAACTCGGCGCGCGCTGGCGCGCGCTTTC 3813
1366 heAspArgGly.....GlyHisGlySerVal... 1374
3814 GGGCAATACGCGATCGGC..... 3831
1375 GlySerLysSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerGlyPh 1391
3832 .....AGTTTCGACATCGGCATC... 3849
1391 eTyrLeuAspGlyValValLysLeuAsnArgPheGluSerAsnValAlaG 1408
3850 .....AGCAGCGCGCGGTTTACGCGCGCATCTTCNACAGCGC 3891
1408 lYlswetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGly 1424
3892 ATCGGAGGCAAAATCGCGCGCGCTGCTGATTAACGCGATTACAGCAG 3941
1425 LeuGlyGlyHis.....IleGluThrCl 1432
3942 ATACGCGCGCGGTTTGGCGGATTCGCGCATGACCGTACATCGCGCA 3991
1433 yMetThrPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeu 1449
3992 CGCGTATTTTCTCCAAAAGCGGATACCGC.....TACGAA 4029
1449 hGlyPheThrAlaAspAsnProGluTyrHisLysSerHisGlyMetLys 1465
4030 AACGTCATATGCGCACCGCGCGCTTTCGTTCAACGCGTACCGN...GC 4076
1466 SerLysSerValAspThrArgSerLe.....TyrArgGluLe 1478

```

```

4077 GGGCATTAAGCAGATATTATTCATTCAAACCGCGCAACACATNTTCATCA 4126
1478 uGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleG 1495
4127 CMCCTTATTNAGCTGTCTTACCGATGCCGCTTCGGGCAAAATCCGA 4176
1495 lProThrLeuLysAlaAlaValArgLysGluPheValAspAsnArg 1511
4177 ACACGCGTCATATAC...GCNGTATTTGGCTCAGATTTTC...GGCAAAAC 4220
1512 ValLysValAsnAsnAspLysAsnPheValAsnAspLeuSerGlyArgA 1528
4221 CGGAGTGGGGAATGGCGGTAAACCGCGGAATCAAGATTTCACGCTGT 4270
1528 gGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer...ThrLeuS 1544
4271 CCNCTCACGCTGCCCGCGCAAGGCGCAACGTGAGCGCAACACAGC 4320
1544 eGlyHisLeuGlyValGly.....TyrSerHisGly 1554
4321 GCGGCGATCAAA.....TTAGGCTACGCGCTGG 4347
1555 AlaGlyValGluSerProTrpAsnAlaValAlaGlyValAsnTrp 1569

```

seq_name: p1r2:D85912

```

seq_documentation_block:
  hypothetical protein ypjA [imported] - Escherichia coli (strain O157:H7, substrain E
  C:Species: Escherichia coli
  C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
  C:Accession: D85912
  R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
  Nature 409, 529-533, 2001
  A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
  A:Reference number: AB5480; MUID:21074935; PMID:11206551
  A:Accession: D85912
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-1528 <STO>
  A:Cross-references: GB:AE005174; NID:g12517083; PIDN:AA657760.1; GSPDB:GN00145; UMG
  A:Experimental source: strain O157:H7, substrain EDL933
  C:Genetics:
  A:Gene: ypjA

```

```

alignment_scores:
  Quality: 354.00      Length: 1667
  Ratio: 0.460        Gaps: 80
  Percent Similarity: 46.191  Percent Identity: 19.376

```

alignment_block:

US-09-303-518D-651 x D85912 ..

Align seg 1/1 to: D85912 from: 1 to: 1528

```

148 AACTACCAATATCTCCGACTTGGCCGAATAAAGCAAGTTTGCAAT 197
124 AsnGlnAspValTyrLysGlyGlyIleSerAsnGlyThrThrIleLysG 140
198 CGGGCGCAAGATTTAGGTNTACACAAAAAAGGAGCTTGTC.... 243
140 yGlyAlaSerArgValGluGlySerAlaAsnGlyThrLeuIleAspG 157
244 ..GGCAATCAATGACAAAAGCCCG...ATGATTGATTTTGTGGGTG 288
157 lYglSerGlnIleValLysValGlnGlyHisAlaAspGlyThrThrIle 173
289 TCGGTACGCGCGTGGCGCATTTGGTG..... 315
174 AsnLysSerLysSerGlnAspValValGlnGlySerLeuAlaThrAsnTh 190
316 .....GGCGATCAATATATTT.....GTGA 334

```

```
190 rThrIleasnlglyIarGlnIyValGluInserThrValGluIthrT 207
      ||| |||||
335 GCGTGGCAGATACGGCGCTATACACAGTT..... 366
      |||||
207 hrThrIleLysasnlglyGluInarValTyrGluSerArgAlaLeu 223
      |||||
367 .....GATTTGGTGGCGAAGGAAAGNATCCCGATCGACCG 404
      |||
224 AspThrThrIleGluGlyThrGlnSerLeuasnSerIysSerThrAl 240
      |||
405 TTTTCTTACCAAAATGTGAAAAGAAATATATTAAGCCATCAATTAC 454
      |||||
240 aLysasnThrGlnIleTyrSerGlyThrGlnIleIleasnThr 256
      |||||
455 ACCCTTACACGGCGATTATATATGCGGCTTGCATTAATTTGTCACA 504
      |||
257 .....Ser 257
505 GATGACAGACCTGTGCAATGACAGT.....GACATGAGGGG 542
      |||||
258 SerSerAspValIleGluValTyrSerGlyValLeuAspValSerG 274
      |||||
543 GAATACCTATTCCGATTAAGAAAATATCCAGCGCTGCCGATCGGCT 592
      |||||
274 yGlyThrAlaThrAsnValThrGln..... 282
593 CAGGACACCACTATTGGCGTTATGATGATACAAACACGGGATTATCC 642
      |||
282 ..... 282
643 TACTCGGCGCATGTGTTAATTGGCGCAATACACAT.....ATGCA 683
      |||||
283 HisAspGlyAla...IleLeuLysThrAsnThrAsnGlyThrThrValSe 298
      |||||
684 GGGTGGGGAATTAATGGCGTANTAGTTTG..... 714
      |||||
298 rGlyThrAsnSerGluGlyAlaPheSerIleHisAsnHisValAlaAspA 315
      |||||
715 .....ACGGCGATGGCGCCATGCCACGACGACGATGAGCCCT 750
      |||||
315 snValLeuLeuGlnasnGlyIyHisLeuAspIleAsnAlaTyrGly... 330
      |||||
751 ATGCCGATTGCAGTGGCGGACGACGCGGTGCCAATGTTTATTATTA 800
      |||||
331 .....SerAlaasnLysThrIleIleIy 338
      |||||
801 TGACAAACAAACATTAATGCTGCTCAAGGAGTTTACAAACCGGCT 850
      |||||
338 snAspLysGlyThrMetSerValIleuThrAsnAlaLysAlaAspAla... 353
      |||||
851 ACCCTATTCCGGGCAAGGAAACGGTTTCCAGCTGATACGCAAGATTGG 900
      |||||
354 .....ThrArgIleAspAsnGlyValMet..... 362
      |||||
901 TTCTACAGATGACATTTTACAGAGCGATACACATACCGTCTNTTTGAACC 950
      |||||
363 .....AspValAlaGlyAsnAlaThrAsnThrIleIle... 373
      |||||
951 GCGCAGTACGACATTTTCTTACATCCAAACAAACAGGTACGGGTA 1000
      |||||
374 .....ThrGlnasnIleasnAsnTyrGlyIleAlaT 387
      |||||
1001 CGGTACAGAAACCAAGAAAGTNTCCATCCAAACCTT.....AAA 1044
      |||||
387 hrGlyThrAsnIleasnSerGlyThrGlnasnIleLysSerGlyLys 403
      |||||
1045 GTACAGACAGTCCGACTGTTTGCAGATCTTGATGAAGTGAATGAAGA 1094
      |||||
404 AlaAspThrThrIleIleSerSerGlySerArgGlnValValGluLysAs 420
      |||||
1095 ACAGATT.....TACCGGCGAGGGGTGTTAATCACTAC 1129
      |||||
```

```
420 pGlyThrAlaIleGlySerAsnIleSerAlaGlyLys..... 432
1130 GTCCAGGTTAAACACGCGTGAACACCTTCTTTATCATTTACGGCAAC 1179
      |||||
433 .....SerLeuIleValIyThrGly 439
      |||||
1180 GGCAAATCATCTTATCAACACATCAACCAAGCGCGGGGCTTTGTA 1229
      |||||
440 Gly.....IleAlaHisGlyValAsnGlnGluThrGlySerAlaLe 453
      |||||
1230 TTTTGAAGTGATTTTACGCTCTCCCTGAAAAACAAACGAAGCTGGCAAG 1279
      |||
453 uValAlaAsn.....ThrG 458
1280 GCGCGGCGGCTTCATACGT.....CAACAGATACCGTT 1314
      |||||
458 LysAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHisThrIle 474
      |||||
1315 ACTTGGAAAGTAAACGCGGTGGCAACGACCGCTGTCCAAATGCGCAA 1364
      |||||
475 Thr.....GlyGlyGluAlaAsnTyrValValLeuGlu...AsnThr 487
      |||||
1365 AGCAGCGCTCAGCTTCAAGCCCAAGGGAACCAAGGCTGATCAGCG 1414
      |||||
487 rGlyGluLeuThrValValAlaLysThrSerAlaLysAsnThrThrIleA 504
      |||||
1415 TGGCGCAGCGTACAGTCATTTTGTGATCAGCAGCAGCAGCATTAAGCMAA 1464
      |||||
504 spAlaGlyGlyLysLeuIleValGlnLysGlnLys..... 516
      |||||
1465 AAACAAACCTTTAGTGAATGGCTTGNTCAGCGGACGAGGTACGTGCA 1514
      |||||
517 .....ThrAspSerThrAr 521
      |||||
1515 ACTGAATCGGATATATCATGTTCAACCCGCAACACTATATTGGCTTTC 1564
      |||||
521 gLeuAsn.....A 524
      |||||
1565 GCGCGCGAGCTTTGATTTAAACGGCGATGCTTTGTTCCACCGTATT 1614
      |||||
524 snGlyValValLeuGluValGlnAspGlyGlyGlnAlaLysHis...Val 539
      |||||
1615 CAAATATCCGATAGAGGCGGATGATGNCNATCAATTAATGCCCAACAC 1664
      |||||
540 GluGlnInserSerGlyAlaLeuIle.....AlaSerThrTh 552
      |||||
1665 ATCCACCGTTACCATATACAGGAATGAAGT.....A 1696
      |||||
552 rSerGlyThrLeuIleGluGlyThrAsnSerTyrGlyAspAlaPheTyrI 569
      |||||
1697 TTACACACAGGAGTGTGAAGATATCAATAGACTTAATTACAGCAAGAA 1746
      |||||
569 leArgAsnSerGluAlaLysAsnValValLeuGlnAsnAlaGlySerLeu 585
      |||||
1747 ATTCGCTAACAGGCTGTGTTGGCAGAGAAGATATACGACCAAGCAACG 1796
      |||||
586 ThrValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnG 602
      |||||
1797 GCGGCTCAACCTT..... 1809
      |||||
602 yLysMetAspValTyrGlyLysAspValGlyThrValLeuasnSerAlaG 619
      |||||
1810 .....GTTTACAGCGCCCGCAGAGAGCCGACCCCNCTGCTT 1848
      |||||
619 LysThrGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLys 635
      |||||
1849 TCCGCGGGAACAAATTAACGCAACATCAGCAACCAAC..... 1890
      |||||
636 GlyLysLysGlnThrValTyrGlyLeuAlaThrGlnLysasnIleGluSe 652
      |||||
1891 .GCGCAATGTTTTCAGCGGCGAGACCGACACCGCCTACATATATT 1939
      |||||
652 rGlyGluGlnIleValAspGlyGlySerThrGluLys.....ThrHisI 667
      |||||
```



```

1192 .....ProteinProAs 1195
3403 ACCACGGCTTCCCGCGCCGCGCCGCGGATTTGCGCAACC 3452
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1195 nProlyspromp..ProlysprompPro..Asp..Prolysp 1210
3453 GCAGCCCAACCGCAACCTCAACCCCAACCGCAGCCTGATNAGC 3502
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1210 OAspProthProgluProthProthProvalProgluysarillet 1227
3503 GTATAGCAATAGCGGTTGAGTATTTCCGACCGCTCAACAGGTT 3552
1227 hrProserThAlaAlaValLeuasnMetAlaAlaThrLeuProLeuVal 1243
3553 TTC.....GCCGTACAGAGCAATTTGACCGCGTGTTC 3587
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1244 pheAspAlaGluLeuasnSerIleArgGluArgLeuasnIleMetIysAl 1260
3588 CGAAGACCGCGCAACGCGTTTGACAGCAGCATCCGCAACCAAC 3637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1260 aserProhIsasnAsnValITPGLyAlaThrTyrasn...ThrArga 1276
3638 ACTACCGTTCCGACGATTTCCGCGCTACCGCAACCAACGACCTGCGC 3687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1276 snAsnValThrThrAspAlaGluAlaGlyPheGluThr 1289
3688 CAATCGGTAGCAAAAACCTC.....GG 3713
1290 LeuthrGlyMetThrValGlyLeaspserProasnAspIleProgluI 1306
3714 CAGCGCGCGCGTCGCGCATCCTGTTTGCAGCAACCGCAACCAACACT 3763
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1306 ylleaThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIleGlyP 1323
3764 TCGACGACGCGCATCGCAACTCGGACGCGCTTCCCGCGCGCTTTC 3813
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1323 heAspArgGly.....GlyHisGlySerVal... 1331
3814 GGGCAATACGCGCATCGC..... 3831
1332 GlySerTyrSerLeuGlyTyrAlaSerTrpGluHisGluSerGlyP 1348
3832 .....AGTTTCGACATCGCGCATC... 3849
1348 eTyThrLeuAspGlyValValLysLeuAsnTrpPheGluSerAsnValAlaG 1365
3850 .....AGCAGCGCGCGGCTTTTACGACGCGCATCTNTCGACAGCGC 3891
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1365 yMetArgPheThrAspGlyAsnTrpAsnLeuthrProTyrAlaSerLeu 1406
3892 ATCGGAGGCAAAATCGCGCGCGCTGCAATTCAGGCATTCAGGACAG 3941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1382 LeuGlyGlyHis.....IleGluThrGlu 1389
3942 ATACCGCGCGCGGTTTTCGCGGATTCGACATCGAACCCTACATCGCGCA 3991
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1389 yMetArgPheThrAspGlyAsnTrpAsnLeuthrProTyrAlaSerLeu 1406
3992 CGCGCTATTTCTCAAAAAGCGGATTACCGC.....TACGAA 4029
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1406 hrcIyPheThrAlaAspAsnProgluTyrHisLeuSerAsnGlyMetGlu 1422
4030 AAGTCGAATTCGACACCGCGCGCTTGGGTTCAACCGTACGCGN...GC 4076
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1423 SerLysSerValAspThrArgSerIle.....TyrArgGluLe 1435
4077 GGGCATTAAAGCAGATTATTCATTAACCGCGCAACATNTCCATCA 4126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1435 uGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleG 1452
4127 CACCTATTNNAGCGTCTCTATACGAGATGCGCTTCGGGCAAGTCCGA 4176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1452 juProthrLeuLysAlaAlaValArgLysGluPheValAspAsnArg 1468
4177 ACACGGGTCAATACC...GCNGTATTTGGCTCAGGATTC...GGCAAAAC 4220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1469 ValLysValAsnAsnAspLysnAspPheValAsnAspLeuSerGlyArgAr 1485
4221 CCGCAGTGGCAATGGGCGGTAAACCGCGCAATCAAAAGGTTTCACGCTGT 4270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4271 CCNTCCAGCGTCCCGCGCCCAAGCNCGCCAATCGAAGCGCAACAGAC 4320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1501 eArgLysHisLeuGlyValGly.....TyrSerHisGly 1511
4321 GCGGCGCATCAA.....TTAGCTACCGCTGG 4347
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1512 AlaGlyValGluSerProThrAsnAlaValAlaGlyValAsnTrp 1526
seq_name: p1r2:T08658

```

```

seq_documentation_block:
  hypothetical protein f949 - Escherichia coli
  C:Species: Escherichia coli
  C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
  C:Accession: T08658
  R:Plunkett, G.
  submitted to the EMBL Data Library, September 1995
  A:Reference number: 216465
  A:Accession: T08658
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-949 <PIU>
  A:Cross-references: EMBL:U36840; NID:g1033110; PID:g1033141
  A:Experimental source: strain k12, substrain MG1655
  C:Genetics:
  A:Map position: 57 min

```

```

alignment_scores:
  Quality: 321.50      Length: 1211
  Ratio: 0.572        Gaps: 50
  Percent Similarity: 46.408      Percent Identity: 19.158

```

```

alignment_block:
  US-09-303-518D-651 x T08658 ..
  Align seg 1/1 to: T08658 From: 1 to: 949

```

```

961 GCACATTTTCTTACATCCAAACACAGGATCGGTACGTAACAGA 1010
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6 GlySerLeuthrValValThrGlySerArgAlaValAspThrIleLeas 22
1011 AACCAAGCAAAAGTNTCCAAATCCAAAGCTTAAATACAGACAGCTCGAC 1060
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 nAlaAsnGlyLysMetAspValTyrGlyLysAspValGlyThrVal.... 37
1061 TGTTTGACGAATCTTTGAATGAATACTGATAAGAACACAGTTTACGGCGCA 1110
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 .....LeuAsnSerAlaGlyThrGlnThrIleTyrAlaSer 49
1111 GGGGGTGTATTCAGTACCGCTCAAGGTTTAAACAAGGTAACGTAACCTTTC 1160
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 .....AlaThrSerAspLysAlaAsnIleLysGlyGlyLysGlnThrVa 64
1161 TTTTATCGATTACGCAACGCAAACTCATCTTATCAACACATCAAC 1210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 L.....TyrGly.....LeuAlaThrGluAlaAsnI 73
1211 AAGCGCGCGCGGTTTGTATTTTAAGGTGATTTTACGGTCCGCGCA 1260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 LeuSerGlyGluGlnIleValAspGlyLysSerThrGluLysThrHis 89
1261 AACCAAC.....GANAAGTGGCAAGCGCGCGCTTCATATCACTGA 1301

```



```

90  IleasnlelylthrglnthrValGlnasntrycilylYsAlaIleasnTh 106
1302 AGACGTACCGTTACTTGAAAGTAACGGCGTGGCAACGACCGCCTGT 1351
106 rAsp..... 107
1352 CCAAAATCGCAAAAGCGACGCTGCAGCTTCAAGCCAAAGGGGAAACCA 1401
108  ....IleValserglyLeuGlnleuGlnIleMetAlaasnlylThrAlaGln 122
1402 GGCCTGATCAGCGTGGCGGACGCTACAGTATTGATTCAGCAGCGACA 1451
123 GlySerIleIle..... 126
1452 CGATAAAGCGCAAAACAAACGCTTTAGTGAATCGGCTTGNTCAGCGCA 1501
127  ....AsnGlyglySerGlnValValasnGlnlylGlyLeu..... 138
1502 GGGGTACGGTGCACGTGAATCCGATATTCAGTTCAACCCCGCAAAACTC 1551
139  ....AlaGlnAsnSerValLeuAsnAsp..... 146
1552 TATTTCGGCTTTCGGCGCGGACGCTTGGATTAAACGGCATTCGCTTTC 1601
147  ....GlylylThrLeuAspValArgIulysGly...se 157
1602 GTTCACCGTATTCAAAATACGATGAGGGCGGATGATTGNMNAICATA 1651
157 rAlaThrnglyIleGlnGlnSerSerIleGlnlyAlaLeuVal..... 170
1652 ATGCCCAACACATCCACCGTTTACCATTAACAGGAATGA..... 1692
171  ....AlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGly 185
1693  ....AGATTACACACAGCGAGTGGTAGAATATCAATAGACTTAA 1733
186 ValAlaIlePheSerIleGlnGlnIleAlaAlaAsnAsnIle..... 198
1734 TTACAGCAAAAGAAATTGCTACACAGGTTGTTGGCGAGAAAGATACGA 1783
199  .... 199
1784 CCAAAACGAAAGCGGCGCTCAACCTTTTACCAGCCCGCGCAAGAC 1833
199  euleuAlaasnlelylValleuThrValGlnSerAspThrSerSerAsp 215
1834 CGCACCCGCTGCTTTCGCGCGGACAAATTTAAACGGCAACATCAGCGCA 1883
216 LysThrGlnValAsnMetGlylArgGluIleValLysThrLysAlaIth 232
1884 AACCAACGCAAACTGTTTTCAGCGGACAGACCGACCGCGCTTACA 1933
232 rAlaIthArgly.....T 236
1934 ATCATTTGGAGAGCGGTTGGTCAAAATGGAAGTATCCACAGAGACAA 1983
236 hrThrleuThrGlylGlyGlnIleValIleGlyValAlaIaasnIleuThr 252
1984 ATCGTGTGGGACACGACTGATCNACCGCAGCTTTAAAGCGGAAATTT 2033
253 ThrIle.....AsnAspGlyGlyIleGlnThrValSerAlaIaasnGlyl 267
2034 CCATATT.....CAGGGCGGCGAGCGGTGATTTCGCCGA 2068
267 uAlaIleLysThrLysIleasnGlnlylGlyThrleuThrValAsnAspA 284
2069 AT.....CTGCGCAAAAGTGAAGGCGATTGNCATTGAGC 2103
284 snGlyLysAlaIthraspIleValGlnasnSerGlyAlaIaIaLeuGlnThr 300
2104 AATCAGCGCCCAAGAGTTTGGTGTGCGACCGCATCAAAAGCCATACAT 2153
::: |||:::
301 SerThrAlaasnlylIleGluIleSerGlyThrHisGlnIlylGlyThrPh 317
2154 CTGTACACGCTTCGACGTGACNGGTCTGCAAAATTTGTGCAANNAANCA 2203
317 e.....Seri 319
2204 TTACGACGATTAAGTGAATGCTTTCATTGACTTAACAGACCTNAGCCGC 2253
319 lAsSerGlyAsnLeuAlaIthrasnMetLeuLeuGlnasnlylGlyAsnLeu 335
2254 ANTGTNAGCTNNCCNATNACGTTNNNTTNAANCTCNCGCCTNCGC 2303
336 LeuValleuAlaGlyThrGlnAlaIarGaspSerThrValGlyLysGlyl 352
2304 ACTNANAGCAATCTTACTGCAATGGCATACAGT..... 2340
352 yAlaMetGlnasnLeuGlyGlnAspSerAlaIthrylValAsnSerGlyG 369
2341 ....TATACGTGAGCCCAACGACCGCACCAAAAGCGACCTTACGCTC 2385
369 lylGlnIlyThrleuGlylArgSerLysAspGluPheGlnAlaIaLeuAlarG 385
2386 GTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAACGGCAACNC 2435
386 AlaGlnAspLeuGlnValAlaGlylGlylThrAlaIaIeValIlyrAlaGlyTh 402
2436 ATCGGNTTCGGGCAATGCTTCATTTAATCTTAAGCAACAGCCGCAACAA 2485
402 rLeuAlaAspAlaSerValSer.....GlylAlaT 412
2486 ACGGACGTGACGCTTTCGCAACAGCTAAGCAACGATTAAGCATTC 2535
412 hrGlySerLeuSerLeuMet...ThrProArgAspAsnValIthProVal 427
2536 GCACCTCAAGGCAATGCTCTCCAGCGCATTAAGCGAGTATTCATTGGA 2585
428 LysLeuGlnGlyAlaValAlarGIlleThrAspSerAlaIthreuthrLeuG 444
2586 AAACAGCCGCTTTCACGCAACTCAGCGGACGAGCAAGANACCATTAAC 2635
444 yAsn.....GlyValAspThrThrleuAla 453
2636 ACTTA.....AAAGACAGCAATGAGACGCTGCGGTGAGGACG 2673
453 spleuthrAlaIaIaSerArglySerValItrp..... 463
2674 GAATTAGCAATTTAAACCTTGACACAGCCACCATT...ACACTCAATT 2720
464  ....LeuAsnSerAsnAsnSerCysAlaGlylThrSerAsnCy 476
2721 CGCCTATCGCACGATGCTGAGGCGGCAAAACCGGAGAGTGTACAGACA 2770
476 sGluThrArgValAsnSerLeuLeuAsnAspGlyAspVal..... 490
2771 CGCGCGCGCGCGCTTCGCGCGCTTCCTATTATCC.....GTTACACG 2814
491  ....TyrleuSerAlaGlnThrAlaIa 498
2815 CCAACTTCGGTAGAATCCGTTTCAACAGCGCTGACGTTAAAGCGCAAT 2864
499 ProAlaIthThrAsnGlyIleTyrAsnThrleuThrThrAsn...GluIe 514
2865 GAACNGTCAGGAACATTCGCTTTATGTCGAGAACCTTCGCTGACGCA 2914
514 userGlySerGlyAsnPherylLeuHsThrAsnValAlaGlySerArgG 531
2915 GCGCAAAATTTGAAGCTGGCGGAAAGTTCCGAAAGNACTTAACCTTGGG 2964
531 LyspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPherylIlePhe 547
2965 GTCAACATATCGGCAACGACCCGTAAGCTTCGATTCATTCAGCTTACT 3014
548 ValGlnAspThrGlyValSerProGlnSerAspAspAlaMetThrleuVal 564
```

```

3015 GGAAGGGAAGACAAACACCGCTGCGAAACCTTAATTCACCTGC 3064
      |||
      |||
564 L.....LysThrIglyIglyAspAlaSerPheThrLeug 576
      |||
3065 AAAAGGAA.....CAGTCGATCGCGCGGCGCTTACCACCTCATC 3108
      |||
576 LysThrIglyIglyPheValAspLeugIlyThrIglyValLeu 591
      |||
3109 CGCAAGACGCGC.....GAGTCCGCGCTGCATATCGGTCAAGAAC 3152
      |||
592 LysSerAspGlyAsnSerAsnThrPasnLeuThrAsnValLys 606
      |||
3153 AGAGCTTCCGACAACTCGCGAAGGACGAAACCAAAACAGCGCGAAA 3202
      |||
3203 AAGACAGCGCGAAGCCTGTAGCGGCTGATGCGCGCGCGGATGCC 3252
      |||
3253 GCGGAAAGACGAAAGCGTTGCGGACCGCGCGCGGCGGCGGGA 3302
      |||
607 ...ProAsnProAspProIleProAsnPro..... 615
3303 AAATGTCGCGCATTTATGACGCGGAGGAAAGAAAAAGCGGTCCAGCGG 3352
      |||
615 ..... 615
3353 ATAAAGACAGCGCNTTGGCGAAACAGCGGAAACCGCGCGGNT 3402
      |||
616 .....LysProAspProLysProAsp 622
3403 ACCACCGCTTCCCGCGCGCGCGCGCGGATTTGCCCAAC 3452
      |||
623 ProLysProAsnProAsnProLys.....ProAsp 633
3453 GAGCGCCCAACCGCAACCTGCAACCGCAACCGCGCGCTGATNACC 3502
      |||
633 OthrProAsnProIleProIleProIleProIleProIleProIlePro 649
3503 GTTATGCGCAATAGCGGTTTGAAGTATTTCCGCCACGCTCAACGCGTT 3552
      |||
650 .....SerThrAlaValAlaValMetAlaIleThrLeuProLeuVal 664
3553 TTC.....GCCGTACAGGACGAAATTCGACCGGCTTGGC 3587
      |||
665 PheAspAlaGlyLeuAsnSerIleArgGlyLeuAsnIleMetLysAl 681
3588 CGAAGACCGCGCAACGCGTTCGACAGCGCAACGCAATCCGNAACCAAC 3637
      |||
681 AsnProIleAsnAsnAsnValIleProIleAlaThrIleAsnIleArgAsnAla 698
3638 ACTACCGTTCCGAAGATTTCCGCGCGCTACCGCAACCAACGACCTGGCG 3687
      |||
698 snValIleThrIleAspAlaGlyAlaGlyPheGlyIleThrIleThrIleMet 714
      |||
3688 CAATCGGATGCGCAAAAC.....CTCGGAGCGCGCGCGCT 3725
      |||
715 ThrValGlyIleAsp**ArgAsnAspIleProGlyIleThrIle 731
      |||
3726 CGGCAATCCGTTTTCGACACCGGACGAAACCAACCTTCGACGAGCGGA 3775
      |||
731 uGlyAlaPheMetGlyTyrSerIleSerIleIleGlyPheAspArgGly 747
      |||
3776 TCGGCAACTCGGACGCGTTCGCCACGCGCGGCTTTCGCGCAATACGCG 3825
      |||
748 .....GlyIleGlySerVal...GlySerTyrSer 756
      |||
3826 ATCGGC..... 3831
      |||
757 LeuGlyGlyTyrAlaSerTyrGlnAsnGlySerGlyPheTyrLeuAspAl 773

```

```

3832 .....AGGTCGACATCGGCATC.....AGCA 3853
      |||
773 yValValLysLeuAsnArgPheLysSerAsnValAlaGlyLysMetSer 790
      |||
3854 CGGCGCGCGGTTTTCGACACCGCGCACTTCGACAGCGCATCGGAGCAAA 3903
      |||
790 erGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyLeuGlyHis 806
      |||
3904 ATCGCGCGCGCGCTGCTGCATTAACGCAATTCAGCGGACGATTCGCGCG 3953
      |||
807 .....IleGlyThrIleMetArgPhe 814
      |||
3954 TTTCGCGGATTCGCGCATCGAACCTGATTCGCGGACGCGCTATTTTCG 4003
      |||
814 rasPGLysAsnThrPasnLeuThrProTyrAlaSerIleThrIglyPheThr 831
      |||
4004 TCCAAAAGCGGATTAACCGCTACGAAAC.....GTCAATATC 4041
      |||
831 laAspAsnProGlyTyrHisLeuSerAsnGlyMetLysSerLysSerVal 847
      |||
4042 GCCACCGCGCGCTTCGCTTCACCGCTACCGN...GCGGCGATTAAGGC 4088
      |||
848 AspThrArgSerIle.....TyrArgGlyLeuGlyAlaThrIle 860
      |||
4089 AGATTATTCATTCAACCGCGCAACATTCATTCACACCTTATTTNA 4138
      |||
860 uSerTyrAsnMetArgLeuGlyAsnGlyMetGlyValGlyProIle 877
      |||
4139 GCCGTCTCTTACCGATCGCGCTTCGCGCAACTCGCAACGCGCTCAAT 4188
      |||
877 ysaAlaAlaValAlaGlyGlyPheValAlaAspAsnArgValLysValAsn 893
      |||
4189 ACCGCGTATTCGCTCAGATTC.....GGCAAAACCGGAGTGGCA 4232
      |||
894 SerAspGlyAsnPheValAsnTyrLeuSerGlyArgArgGlyIleTyrG 910
      |||
4233 ATGGCGGCTAAACCGCGCAATCAAGTTTCACGCTTCCTCCNTCCAGCGT 4282
      |||
910 nAlaGlyIleLysAlaSerPheSer...ThrLeuSerGlyHisLeuG 926
      |||
4283 CCGCGCGCAAGCGCGCACTGGAACGCGCAACAGCGCGCATCAAA 4332
      |||
926 lValGly.....TyrSerHisSerAlaGlyValGly 936
      |||
4333 .....TTAGCTTACCGCTGC 4347
      |||
937 SerProThrPasnAlaValAlaGlyValAsnTyr 947
      |||
seq_name: p1r2:E95206
seq_documentation_block:
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:accession: E95206
R:Telletlin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umeyam, L.A.; White, E.K.; Salzborg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; M01D:2157209; PMID:11463916
A:Accession: E95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4776 <KUD>
A:Cross-references: GB:AE005672; PIDN:AKK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI772
alignment_scores:

```

Quality: 298.50 Length: 1132
Ratio: 0.464 Gaps: 34
Percent Similarity: 56.802 Percent Identity: 20.053

Alignment block:

US-09-303-518d-651 x E95206 ..

Align seg 1/1 to: E95206 from: 1 to: 4776

```

1100 TTATCGCGGAG...GGGGTGAATACAGTACCTGCA...CG 1137
|||||:|||||:|||||:|||||:|||||:|||||:
40 PheThrAsnGluSerAlaValLeuGluThrValGluLysThrAspAl 56
1138 TTAACACAGCGTAAACCTTTCTTTATGATACGCGACGCGCAACT 1187
:|||||:|||||:|||||:|||||:|||||:|||||:
56 AleuAlaThrAsnAspThrValValLeuGluThrIleSerThrSerAsn 73
1188 CATCTTATCAACACACA.....TCACCAAGCGC 1216
|||||:|||||:|||||:|||||:|||||:|||||:
73 eTAlaSerSerThrSerLeuSerAlaSerGluSerAlaSerThrSerAla 89
1217 CGGGCGGTTGTATTTTGAAGGATTTACGGTCTCGCTGAAACAAAC 1266
:|||||:|||||:|||||:|||||:|||||:|||||:
90 SerGluSerAlaSerThrSerAlaSerThrSerAlaSerThrSerAla 106
1267 GAACGTGGCAAGCGCGCGCGCTCATATCACTGAAG...ACAGTACCGT 1313
:|||||:|||||:|||||:|||||:|||||:|||||:
106 rGluSerAlaSerThrSerAlaSerThrSerIleSerAlaSerThrIv 123
1314 TACTGGAAAGTAAAGCGCGTGGCAAGACCGCTGTCACAAATGCGCA 1363
:|||||:|||||:|||||:|||||:|||||:|||||:
123 aIValGluSerGluThrAlaAlaIleThrGluAlaThrAlaLysVal 139
1364 AAGCAGCGTGCAGCTTCAGCCAAAGGAAACCAAGCGCTCATGAGC 1413
:|||||:|||||:|||||:|||||:|||||:|||||:
140 Glu...GluAspArgLysLysProAlaSerAspLysValAlaSerValTh 155
1414 GTGGCGCAGCGTACAGTATTTGGATCAGCAGCAGCATTAAGGCA 1463
:|||||:|||||:|||||:|||||:|||||:|||||:
155 rAsnValAsnLeuGlnSerThrAlaLysArgLysArgSerValAspS 172
1464 AAAAAACCTTTAGTAAATCGGCTGNTCAAGCGCGCAGGGGTACGGTGC 1513
:|||||:|||||:|||||:|||||:|||||:|||||:
172 eTleGluGlnLeuLeuAlaSerIleLysAsnAlaAlaValPheSerGly 188
1514 AACTGAATGCCGATATATCAAGTTCACCCGACAAACTCTATTTCGGC... 1560
|||||:|||||:|||||:|||||:|||||:|||||:
189 Asn.....ThrIleValAsnGlyAlaProAlaIleAsnAlaSerLeuAs 203
1561 .....TTTCGGCGCGGCGAGCTTGATTTAA 1585
203 nIleAlaLysSerGluThrLysValThrGlyGluGlyValAlaSpSerV 220
1586 ACGGCATTCGCTTTCCTCCACCGTATTCAAAATACCGATGAAGGGCG 1635
:|||||:|||||:|||||:|||||:|||||:|||||:
220 aLTYrArgValProIleThrLysLeuLysValThrAsnAspGlySer 236
1636 ATGAT..... 1641
237 LysLeuThrPheThrValThrValThrValAsnProLysThrAsnAs 253
1642 .....GNCATCATATATGCA 1657
253 pLeuGlyAsnIleSerSerMetAlaProGlyTYrSerIleThrLysSerG 270
1658 CAACACATTCACCGTTACATTAACAGGAATGAAGTATTAACACACCG 1707
:|||||:|||||:|||||:|||||:|||||:|||||:
270 LYrThrSerThrGlnThrMetLeuThrLeuGlySerAspLeuGlyLysPro 286
1708 AGTGTAAGATATATCAATAGACTTAATTAACAGCAAAATTCGCTACAA 1757
|||||:|||||:|||||:|||||:|||||:|||||:
287 SerGlyValLys.....AsnTYr..... 292

```

```

1758 CGGTGTTGGCGGAGAAAGTATACGACCAAAACGACGGCGCTCAAC 1807
293 .....IleThrAspLysAsnGlyArgGlnValL 302
1808 TTGTTACAGCGCGCGCAGAACCGCACCCNCGCTCTTCGCGCGGA 1857
|||:|||||:|||||:|||||:|||||:|||||:
302 euSerTYrAsnThrSer.....ThrMetThrThrGlnGlySer 314
1858 ACAATTTTAAACGGCACATCAGCAACCAACGCAAACTGTTTTCAG 1907
|||||:|||||:|||||:|||||:|||||:|||||:
315 GlyTYrThrThrPGLysAsnGlyAlaGlnMetAsnGly.....PhePheAl 329
1908 CGGACAGACCGACACCGCGCAGCTTACATCATTTAGAACGCGGTGCA 1957
:|||||:|||||:|||||:|||||:|||||:|||||:
329 aLysLys.....GlyTYrGly...LeuThrSerSerThrPhe 340
1958 AAATGAGAGTATCCCA...CAAGAGAAATCGTGGGCAACAGACTGG 2004
:|||||:|||||:|||||:|||||:|||||:|||||:
341 .....ValProIleThrGlyThr.....AspThrSerPhe 350
2005 ATCNAACGACGTTTAAACGGGAA.....AATTCGA 2036
:|||||:|||||:|||||:|||||:|||||:|||||:
351 ThrPheThrProTYrAlaAlaArgThrAspArgIleGlyIleAsnTYrPh 367
2037 TATTCAGGCGCGCAGCGCGTGTATTCGCCAATGTTGCCAAAGTGAAG 2086
|||||:|||||:|||||:|||||:|||||:|||||:
367 eAsnGlyGlyGlyLysValValGluSerSerThrThrSerGln..... 381
2087 GCGATTGNCATTTAGCAATCAGCCCAAGCAGTTTGGTTCGCGCAG 2136
|||||:|||||:|||||:|||||:|||||:|||||:
382 .....SerLeuSerGlnSerLysSerLeuSerValSerAlaSerGln 395
2137 CATCAAGCCATACATCTGACAGTTCGACTGACGNCGTGCAAA 2186
:|||||:|||||:|||||:|||||:|||||:|||||:
396 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThr 412
2187 TTGTGTGAANAANCAATACCGCAGTAAAGTATGTTGCTTACACTA 2236
:|||||:|||||:|||||:|||||:|||||:|||||:
412 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSer..... 426
2237 AGACNGACNTNAGCGCANTGTNAGNCTNACATNACGNTNTTAAAN 2286
|||||:|||||:|||||:|||||:|||||:|||||:
427 ..ThrSerAlaSerValSerAlaSerThrSerAlaSerAlaSer 442
2287 CTCNCGGGGTCGNNCATTNANGCAATCTAGTGAATGGCATAC 2336
:|||||:|||||:|||||:|||||:|||||:|||||:
443 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerTh 459
2337 ACGTTATACAGTACGCCACACGCCACCAAAACGCGCAACTTACGCTCG 2386
:|||||:|||||:|||||:|||||:|||||:|||||:
459 rSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThS 476
2387 TGGGCAATGCCCAAGCAACTTATCAAGCCACATTAACGGAACNCA 2436
:|||||:|||||:|||||:|||||:|||||:|||||:
476 eAlaSerAlaSerAlaSer...ThrSerAlaSerGluSerAlaSerThr 491
2437 TCGGNTTCGGCAATGCTTATTAATCAAGAACACCGCGCAAAA 2486
|||||:|||||:|||||:|||||:|||||:|||||:
492 SerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThSe 508
2487 GCGAGCTGTGACGCTTCGACACAGCGTAAAGCAACGTAAGCAATTCG 2536
:|||||:|||||:|||||:|||||:|||||:|||||:
508 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerA 525
2537 CACTCAAGGCAATGTCCTTAGCCGATTAAGCACTATTCATTTTGA 2586
|||:|||||:|||||:|||||:|||||:|||||:|||||:
525 Ia...SerGlySerAlaSerThrSerThrSerAlaSerAlaSerThrSer 540
2587 AACAGCGCTTACCGGACACTCAAGCAGCGCACAGGAGAAACGATTCGA 2636
|||||:|||||:|||||:|||||:|||||:|||||:
541 AlaSerAla...SerAlaSerThrSerAlaSerAlaSerAlaSerIleSe 556
2637 CTTAAAGACAGCAATGAGCGCTGCGCTCAGGACGAGAAATTAAGCAATT 2686

```


A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90674
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1327 <HAY>
A:Cross-References: GB:BA000007; PIDN:BA833785.1; PID:q13359819; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS0362

alignment_scores:

Quality:	285.50	Length:	1542
Ratio:	0.418	Gaps:	71
Percent Similarity:	44.293	Percent Identity:	18.872

alignment_block:

US-09-303-518d-651 x B90674 ..

Align seg 1/1 to: B90674 from: 1 to: 1327

```
364 GTTGATTGTCGCGAAGAGNATCCGATCAGCACCCTTTTCTTA 413
      ::::::::::: ||::: ||
59 ILAspPheGlyLysGlnSerIleAsp.....Ty 70
      ::::::::::: ||::: ||
414 CCAAAATTGTAAGAAATATTATTAG.....CTGACA 448
      ::::::::::: ||::: ||
70 rArgIleThrValThrAspAsnIleAsnLeuValIleAsnAlaThrAspT 87
      ::::::::::: ||::: ||
449 ATTACACACCTTTCACAGCGGATTATACGCGCGTTGCATTAATT 498
      ::::::::::: ||::: ||
87 hIserArgPro..... 90
      ::::::::::: ||::: ||
499 GTCCACAGATGCAGAACCTGCAATGACAGTGCATGAGGAGGAATAC 548
      ::::::::::: ||::: ||
90 ..... 90
      ::::::::::: ||::: ||
549 CTTATCCGATTAAGAAATATCCGAGCGTCCGATCGGCTCAGAG 598
      ::::::::::: ||::: ||
91 .....ArgLeuThrIleAlaIleSerGly. 97
      ::::::::::: ||::: ||
599 ACCACTATTGGCGTTATGATGATGACAAACAGCGGATTATCCACTCC 648
      ::::::::::: ||::: ||
98 .....GlyGlyLeuAspIleThr 103
      ::::::::::: ||::: ||
649 GCGCGATGGTTA...ATTGGCGGC.....AATACACATATGACAGGTG 689
      ::::::::::: ||::: ||
104 GlyGlyLysValThrIleAsnGlyProLeuAsnPheLeuLeuGlyTh 120
      ::::::::::: ||::: ||
690 GGGA.....AATAATGGCGTANTTATGTTGAGCGCGAT... 723
      ::::::::::: ||::: ||
120 rGlyPheLeuAsnValSerAsnAlaGlySerGluLeuTyraIleAspAsp 137
      ::::::::::: ||::: ||
724 .....GTGGCGCATGCCAACGACGACTATGGCCCTATG 753
      ::::::::::: ||::: ||
137 euTyrgLuserAsnSerGlyMetArgHisAspArgGlyTyrrPheAsnVal 153
      ::::::::::: ||::: ||
754 CCGATTGACGGTGGCGACAGCGCGGTTCGCCAATGTTTATTTAT... 801
      ::::::::::: ||::: ||
154 SerAsnGlyGlyLysIleHisValLysGlyThrSerArgLeuThrTyLe 170
      ::::::::::: ||::: ||
802 .....GACAAAACAACAATAAATGCGCTGC 826
      ::::::::::: ||::: ||
170 uGlnIleLysnValSerGlyGluGlySerGlnValAsnSerGluThrPhe 187
      ::::::::::: ||::: ||
827 TCACGGAGGATTTTACAAC.....GGCTACCTTATTCGCGAGGGA 870
      ::::::::::: ||::: ||
187 hemeGlyValTyrgLysertyrGlyAsnGlnTyrrLeuSerValAsn 203
      ::::::::::: ||::: ||
871 AACGGTTTCCAGCTG...ATACGCAAGAT.....TGGTTCTAGGA 908
      ::::::::::: ||::: ||
204 AsnGlyGlyGluValAsnAlaArgLysGlnIleSerLeuGlyTyrrTyAs 220
      ::::::::::: ||::: ||

909 TCACATTTACAGAGCGGATACACATACCGCTCTNTTTGAA.....CCGC 952
      ::::::::::: ||::: ||
220 pGlnVal.....SerAspThrThrLeuAlaValSerGluGlyLysI 235
      ::::::::::: ||::: ||
953 GCAGTAACGACATTTTCCCTTACATCCCAACAACAAC..... 990
      ::::::::::: ||::: ||
235 lSerAlaProThrIleSerLeuSerThrAsnSerGlnLeuAlaLeuGly 251
      ::::::::::: ||::: ||
991 ...GGTACCGGTACGTAACAGAAACCAACGAAAGCTTCCATTCAAA 1037
      ::::::::::: ||::: ||
252 AlagInGluGlySerAlaIleAlaLysAlaIleGlyIleIleAspAlaGly 268
      ::::::::::: ||::: ||
1038 GCTTAAGTACAGACATCCGACTGTTGACGA.....TCTTGAATG 1081
      ::::::::::: ||::: ||
268 sIleGluPheValITrPalAlaLysThrSerGluLysLysIleThrLeuAsn 285
      ::::::::::: ||::: ||
1082 AAACGTATAAAGAACCGTTTACGGGAGCGGCGGTTAATCAGTACCGCT 1131
      ::::::::::: ||::: ||
285 lThrAspLysAspAlaThrIleSerAlaAspIleVal..... 297
      ::::::::::: ||::: ||
1132 CCAAGGTTAAACACAGCGTAAACCTTCTTTATCGATTACGGCAACGG 1181
      ::::::::::: ||::: ||
298 .....SerGlySerGluGlyLeuGlyTyrrIleAsnAlaLeuAsnG 311
      ::::::::::: ||::: ||
1182 CAAACTCATCTTATCAACACATCAACACAGCGCGCGGCTTTGATTT 1231
      ::::::::::: ||::: ||
311 YThrThrTyrrLeuThrGlyAspAsnSerAlaPheSerGlyLysValLys 328
      ::::::::::: ||::: ||
1232 TTGAAGTGATTTTACGGTCTCGCTGAAACAAAC.....GAA 1269
      ::::::::::: ||::: ||
328 lGluGlnAsnGlyAlaLeuGlyIleThrGlnAsnIleGlyThrAlaGlu 344
      ::::::::::: ||::: ||
1270 ACGTGCAGAGCGCGGCGTTCATATCAGTGAACAGTACCGTTACTTG 1319
      ::::::::::: ||::: ||
345 lIleAsnAsnArgGlyLysLeuHisLeuLysAlaAspSerMetThrPh 361
      ::::::::::: ||::: ||
1320 GAAAGTAAACGCGGTGGCAACGACCGGCTGTCCAAATGGCAAGGCA 1369
      ::::::::::: ||::: ||
361 e...AlaAsnLysIleSerGlyAsnGlyThrIleSerIleAspSerGlyT 377
      ::::::::::: ||::: ||
1370 CGCTGCACGTTTCAAGCCAAAGGGGAAAC.....CAAGGCTCGATC 1410
      ::::::::::: ||::: ||
377 hr.....ValGluLeuThrGlyAsnAsnTyraIlePheSerGlyTyrrIle 391
      ::::::::::: ||::: ||
1411 AGCGTGGGCGAGGTACAGTCATTTTGGATCAGACGACGACGATAAG 1460
      ::::::::::: ||::: ||
392 AspValAlaSerGlyAlaValAlaValIleSerGluAspLysAsnIleG 408
      ::::::::::: ||::: ||
1461 CAAAACCAACGCCCTTATGTAATCGGCTTG..... 1491
      ::::::::::: ||::: ||
408 YArgAlaGluLeuAspValAspGlyLysLeuGlnIleAsnAlaAsnLys 425
      ::::::::::: ||::: ||
1492 .....NTCAGCGCGAGGGGTACGTCAGTCAATG 1521
      ::::::::::: ||::: ||
425 sPTrrValIleAspAsnAspLeuGluGlyArgGlyIleValGluIleAsn 441
      ::::::::::: ||::: ||
1522 GCCGATTAAT...CAGTCAACCCCGACAACTCATTTTC..... 1557
      ::::::::::: ||::: ||
442 MetGlyAsnHisGluPheSerPheAspGluPheAlaTyrrThrAspTrph 458
      ::::::::::: ||::: ||
1557 ..... 1557
      ::::::::::: ||::: ||
458 eGlnGlySerLeuAlaPheGlnAsnThrThrPheAsnLeuGluLysAsn 475
      ::::::::::: ||::: ||
1557 ..... 1557
      ::::::::::: ||::: ||
475 lGluPheLeuGlnLysGlyIleThrAlaGlyGlnGlySerLeuVal 491
      ::::::::::: ||::: ||
1558 .....GGCTTGGCGGCGG 1571
      ::::::::::: ||::: ||
492 ThrValGlyLysGlyAlaHisSerIleSerThrLeuGlyPheSerGlyAl 508
      ::::::::::: ||::: ||
1572 ACGTTTGATTTAAACGGGCGATTCGCTTTCGTTCCACCGTATTCAAAATA 1621
      ::::::::::: ||::: ||
```

508 yThrValasphegIyala.....L 515
1622 CCGATGAGGGCGGATGTCNCAATATATGCCACAAACATCCACC 1671
515 eutThrAlaGlyAlaGlnmetThrGluGlyThrValAsnValSerIythr 531
1672 GTTACCATTAACGGAATGAAAGTATTAACACACCGAGTGTGAATAT 1721
532 LeuAspLeuArgGly...GluGlyValIleGlnValSerAspSerAspVa 547
1722 CAATGACTTAATTAACGCAAAAGAAATGCTTACACCGTGTGGTGGCG 1771
547 lValArg...SerValSerArgAspIle...AspSerAlaLeuSerL 561
1772 AGAAAGATACGACCAAAACGAGCGGCTGACACTTGTTCACGACC 1821
561 eutThrGluValAspAspGlyAsnSerThrIleLysLeuValAspAlaGln 577
1822 GCCGAGAAAGACCGCACCCNGCTGCTTCCGGCGGAAACAATTTA.... 1866
578 GlYAlaGln.....ValLeuGlyAspAlaGlyAsnLeuGlnLe 590
1867AACGCAACATC.....ACGC 1882
590 uGlnAspLysAsnGlyGlnIleLeuSerSerSerAlaGlnArgAspIleG 607
1883 AAACAACGGCAACTGTTTTCAGCGC.....AGCCGACA 1920
607 lInGlnAsnGlyGlnLysAlaAlaValGlyThrYrAspIyArgLeuThr 623
1921 CGCAGCGCTACAAAT.....CATTTAGAACGGGTGTGTCAAAT 1961
624 SerGlyValAsnAsnAspGlyLeuIyThrIleGlyIyThrGlyLeuThrGlnLe 640
1962 GGAAGGTATCCCAAGAGAAATCGTGGCAACAACGATGATNACC 2011
640 uAsp..... 641
2012 GCACGTTTAAACGGAAATTTCCATATTACAGGGGCGGCGGTGAT 2061
642LeuHisAlaThrAspSerAspAlaLeuVal 651
2062 TCCCGCAATGTCCTCAAAAGTGGAGGCGATGNCATTTGAGC..... 2103
652 LeuSerSerAsnGlyLysSerGlnAsnAlaAlaAspLeuSerAlaLysI 668
2104AATCAGCCCAAGCAGTTTGGTGTGCACCGCATCAACGCATA 2149
668 eThrGlySerGlyAspLeuAlaPheSerSerGlnLysGlyIlnThrValS 685
2150 CAATCTGTACAGT...TCGACGTGACNGTGTGCAAAATGTGTGCA 2196
685 erLeuSerIlnLysAspAsnAspIyThrGlyValIlnThrAspLeuArgSer 701
2197 NAAANC.....ATTACGACGATTAAGTATGCTTCATGTACTAGAC 2240
702 GlyThrLeuLeuAsnAsnAspAsnValLeuGlyAsn..... 714
2241 NGACNTNACGGCANTGTNAGNCTNNCAATNACGNTNNTTNAANCNTCN 2290
715ThrHisGluLeuArgLeuAlaIleGluThrGluLeuAspMetA 729
2291 CNGCGCTCNCNCACTNNAANGCAATCTTAGCAAAATGGCATACAGT 2340
729 snGlyHisSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeu 745
2341 TATACAGTACGCCAACGCCCAAAAGCAACCTTAGCCTGTGTGG 2390
746LeuSerLeuAsnGly 750
2391 CAATGCCCAAGCAACATTTAATCAAGCCATTTAAAGGCAACNCATCGG 2440
750 yGlySerLeuThrValThrAsnGlyIyThrSerThrGlySerLeuThrG 767
2441 NTTCGGGCATGCTTTCATTTAATCTAGCAACACGCCGCCCAAAACGGC 2490
767 LysGlyGlnLeuAsnIle.....GlnGlyGly 776
2491 AGCTGACGCTTCCGACCAACGTTAAGCAACGTAAGCCATTCGCAC 2540
777 ThrLeuAspIleAlaGlyAsp.....AsnSerAsnLe 787
2541 CAACGCAATGTCTCCCTAGCCGATNAGCA.....GTAATTCATT 2581
787 uThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerHisA 804
2582 TTGAAAACAGCGCTTTACCGACACACTCAGCGGCGACGAAGANACAGA 2631
804 lalngly..... 806
2632 TTACACTTAAAGACAGCAATGACGTCGCCGTCAGACGGAATTAG 2681
807LeuG 808
2682 CAATTTAAACCTTGAC...AAGCCACATTAACCTCAATTCGCTATC 2728
808 ySerAlaAsnValGlnAsnAsnGlyThrLeuAlaLeuAsnAsnSerAlaG 825
2729 GCCACGATGCTGCGAGGCGCAACCGGAGAGNGTGTACAGACGCCGCC 2778
825 lLysAlaGlnAlaAlaSerValAsnIyThrAlaLeuGlyGlnLeuThr 841
2779 CGCGTTCGGCGCTTCCTATTATCCGTTACACCCCACTTCGTTACA 2828
842 AsnAsnGly.....ThrLeuMetThrGlyMetSerGlyG 853
2829 ATCCCGTTTACACAGCTGACGGTAAACGGCAATTTGAC...NGCAG 2875
853 nGlnAlaGlyAsnValLeuValAlaLysGlyAsnIyThrHisGlyAsnAsnG 870
2876 GAACATTCGCTTATGTCGAACTCTCGCTACCGAAGC.....GAC 2919
870 lGlnLeuValMetAsnThrValLeuAsnGlyAspSerValIlnThrAsp 886
2920 AAATTTGAAGTGGCGGAAATTCGGAAGNACTTACACCTTGGCGGTCA 2969
887 LysLeuValAlaGlnGlyAspThrSerGlyThrAlaValThrValAs 903
2970 CAAT.....ACCGCAACGAAACCGCTGACCTGATCAATGACGG 3010
903 nAsnAlaGlyIyThrGlyAlaLysThrLeuAsnGlyIleGluLeuIleH 920
3011 TAGTGAAGGAAAGACACAAACCGCTGCCGAAACCTTAATTTCACC 3060
920 lValAspGlyLysSerGlnGlyIlnPhe..... 929
3061 CTGCAAAACGAAACGTCGATGCGCGGCGTGGCGTTACCACTCATCCG 3110
930 ValGlnAlaGlyArgIleValAlaGlyAlaIyThrAspIyThrIleuAlaAr 946
3111 CAAAGACGGCGAGTTCGCCCTGCATTAATCCGGTCAAGAAACAAGACTTT 3160
946 gGlyGlnGly..... 949
3161 CCGACAAACTGCGCAAGCGACAGCAACCAAAAAACAGGGGAAAAAGACAC 3210
949 949
3211 GCGCAAGCCTTGAAGCGGCTGATGCGCGGCGGCGATCCGCGAATA 3260
950 AlaAsnSerGlyAsnIyPylLeuThrSerGlySerIySerIyPylGlu... 965
3261 GACAGAAAGCGTTGCCGAACCGCGCGGCGGAGGGAATAATGTCG 3310
965 965

```

4121 CC.....ATCAACNCTTATTNAGCGT 4143
      ::||| :::::
1186 lysergInglYlThrArgSnglUTrpTyValGlnPrzGlnAlaGlnVal 1202
4144 TCCTATTCGCATGCC..... 4158
      :::::
1203 ThrTrpMetGlnValLysAlaAspLysHisArgGluSerAsnGlyThrLe 1219
4159 .....GCTTCGGGCAAGTCGACACGGCGTCATATACCGCNG 4195
      :::::||||| :::::
1219 uValHisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysT 1236
4196 TATTGGCTCAG.....GATTTCGGCAAAACCCGC..... 4224
      :::::
1236 hrTrpLeuLysSerHisHisLysMetLAspArgGlyLysSerArgGluPhe 1252
4225 .....AGTGGGAATGGGGCGTAAAGCCGCAATCAAGGTTT 4262
1253 GlnProPheValGlnValAsnTrpLeuHisAsnSer.....LysAspRh 1267
4263 CACGCTGTCNCCTCCACGCTGCCGCGCAAGNCCGCAACTGGAAGCGC 4312
      :::::||||| :::::
1267 eSerTrpSerMetAspGlyValSerVal.....ThrGlnAspGlyAlaA 1282
4313 AACACAGCGCGGCGCATCAATATTAGC 4338
1282 rgAsnLeuAlaGlnLeuLysThrGly 1290

seq_name: p1r2:E85524

seq_documentation_block:
probable beta-barrel outs membrane protein Z0402 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85524
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: E85524
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <STO>
A:Cross-references: GB:AE005174; NID:912513130; PIDN:AA654657.1; GSPDB:GN00145; UWGP
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0402

alignment_scores:
      Quality: 285.50      Length: 1542
      Ratio: 0.418      Gaps: 71
Percent Similarity: 44.293      Percent Identity: 18.872

alignment_block:
US-09-303-518D-651 x E85524

Align seg 1/1 to: E85524 from: 1 to: 1349

364 GTTGATTTTGGTCGGAAGAGAAAGNAATCCGATCGACACCGTTTCTCTTA 413
      ::||| ||| :::::
81 IleAspPheGlyGlyGluAsnGlnSerIleAsp.....Ty 92
414 CCAATTTGTGAAGAATAATTAATAG.....CCTGACA 448
      ::||| :::::
92 TArgIleThrValThrAspAsnAlaAsnLeuValIleAsnAlaThrAspT 109
449 ATTACACCCCTTACAAAGCGCATTCATCATATGCGCGTTTGCAATAAATT 498
      ::||| :::::
109 hrSerArgPro..... 112
499 GTCCACAGATGCAGAACCTGTGCAAAATGACGAGTGCATGAGGCGGAATAC 548

```



```

1962 GGAAGTATCCACACAGAGAAATCGTGTGGACAAAGACTGGATGTCNAC 2011
      ::::
662 uasp..... 663
2012 GCACGTTTAAAGCGGAAAATTTCCATTTTCAGCGCGGAGCGGTGATT 2061
      ::::
664 .....LeuHisAlaThrAspSerAlaLeuVal 673
2062 TCCCGCAATGTGCAAGTGAAGCGGATTCGATTCAGC..... 2103
      ::::
674 LeuSerSerAsnGlySerGluAsnAlaAlaAspLeuSerAlaLys11 690
2104 ....AATCAGCCCAAGCAAGTTTGTGTGCGACCGCATCAAGGCCTA 2149
      ::::
690 eThrGlySerGlyAspLeuAlaPheSerSerGlnGlyGlnThrValS 707
2150 CAATCTGTACAGT...TCGGACTGACAGCGTGTGCAAAATGTGTGAA 2196
      ::::
707 eThrLeuSerAsnLysAspAsnAspTyrThrGlyValThrAspLeuTrgSer 723
2197 NAANAC.....ATTACGACGATTAAGTATTCCTTCATTCAGCTAAGAC 2240
      ::::
724 GlyThrLeuLeuLeuAsnAsnAspAsnValLeuGlyAsn..... 736
2241 NCACNTNAGCGGCANTGTNAGCNTNCCNATNAGCNTNNTTNAANCTCN 2290
      ::::
737 .....ThrHisGluLeuArgLeuAlaIaGluThrGluLeuAspMetA 751
2291 CNGGCGNTGCNNCAGCTNANAGCAATCTTAAGCAATGGCATGACAGCT 2340
      ::::
751 snGlyHisSerGlnThrValGlyThrLeuAsnGlySerAlaAspSerLeu 767
2341 TATACAGTCAGCCACACGCCACCAAAAGCGCACTTAAGCGCTGTGG 2390
      ::::
768 .....LeuSerLeuAsnGly 772
2391 CATGCCCCAGCAACATTTAATCAAGCCACATTTAAAGCGCAACNCATCG 2440
      ::::
772 yGlySerLeuThrValThrAsnGlyGlyThrSerThrGlySerLeuThrG 789
2441 NTTCGGGATGCTTCATTTATTAACAGCAACAGCCGCGCAAAAGCGC 2490
      ::::
789 ySerGlyGluLeuAsnIle.....GlnGlyGly 798
2491 AGTCTGACGCTTTCGACACAGCCTAAGCAAAAGCTAAGCCATTCGCACT 2540
      ::::
799 ThrLeuAspIleAlaGlyAsp.....AsnSerAsnIle 809
2541 CAAGCGCAATGTCTCCCTAAGCGGATTAAGCA.....GTATTCATTT 2581
      ::::
809 uThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerHisA 826
2582 TTGAAMAAGCGCGCTTACCGGACAACTGACGGGAGCAAGANACAGCA 2631
      ::::
826 IaGlnGly..... 828
2632 TTACACTTAAAGACAGCGAATGAGCGCTGCGACAGCAAGATTAGG 2681
      ::::
829 .....LeuG1 830
2682 CAATTTAAACCTTGAC...AAGCGACCATTAACACTCAATCCGCCCTATC 2728
      ::::
830 ySerAlaAsnValGluAsnAsnGlyThrLeuAlaLeuAsnAsnSerAlaG 847
2729 GCGACGATGTGACAGCGCGCAACCGGACAGNGTGTGACAGACACCGCGC 2778
      ::::
847 IuLysArgAlaAlaAlaSerValAsnTyrAlaLeuGlyGlyAsnLeuThr 863
2779 CGCGATTGCGCGCGTCCCTATTATTCGCTTACACCGCAACTTCGGTGA 2828
      ::::
864 AsnAsnGly.....ThrLeuMetThrGlyMetSerGlyG1 875

```

```

2829 ATCCCGTTTCAACACGCTGACGTTAAAGCGCAATTTGAC...NETCAAG 2875
      ::::
875 nGlnAlaGlyAsnValLeuValValGlyGlyAsnTyrHisGlyAsnAsnG 892
2876 GAACATTCGCGCTTTATGCGGAACCTTCGGCTACCGAAGC.....GAC 2919
      ::::
892 yGlnLeuValMetAsnThrValLeuAsnGlyAspAspSerValThrAsp 908
2920 AAATTTGAAGCTGCGGAAAGTTCCGACAGCACTTAACACTTGGCGGTCAA 2969
      ::::
909 LysLeuValValGluGlyAspThrSerGlyThrThrAlaValThrValAS 925
2970 CAAT.....ACCGCAACGAACCGGTAAAGCTGATCAATTTGACGG 3010
      ::::
925 nasnAlaGlyGlyThrGlyAlaLysThrLeuAsnGlyIleGluLeuValLeu 942
3011 TAGTGAAGGAAAGACACAAACCGCTGTCGAAACCTTAATTTTCAAC 3060
      ::::
942 lSValAspGlyLysSerGluGlyLysPhe..... 951
3061 CTGCAAAAGCAACAGCTGATCCGCGCGCGGTACCAACTATCAGCG 3110
      ::::
952 ValGlnAlaGlyArgGlyValAlaGlyAlaTyrAspTyrThrLeuAlaAr 968
3111 CAAGAGCGCGAGTTCGCCCTGCATATCCGCTCAAAAGACAGAGCTTT 3160
      ::::
968 gGlyGlnGly..... 971
3161 CCGCAAACTGCGCAAGGACAGAACCAAAAGCGGGAAGAGACAC 3210
      ::::
971 ..... 971
3211 GCGCAAAAGCTTGAACGCGCTGATTCGCGCGCGCGATCCCGGCAAAA 3260
      ::::
972 AlaAsnSerGlyAsnTyrPylLeuThrSerGlySerAspSerProGlu.. 987
3261 GACAGAAAGGTTTGCAGACCGCGCGCGCGGAGGAGGAAATGTGCG 3310
      ::::
987 ..... 987
3311 GCATTTATGAGGCGGAGAGAGAGAAAAACGGGTGACAGCGGATTAAGAC 3360
      ::::
988 .....LeuGlnProGluProAsp 993
3361 AGCGCNTTGGCGAAGACGCGGAGGAGAAACCGCGCGGNTACCAACCGC 3410
      ::::
994 .....PrometProAsnPr 998
3411 CTTCGCCCGCGCGCGCGCGCGCGGATTTGCGCAACGCGACGCGCC 3460
      ::::
998 ogLupProAsn.....ProAsnProGluProA 1007
3461 AACCGCAACTCAACCCCAACGCGACGCGACCTG..... 3495
      ::::
1007 snProAsnProThrProThrProGlyProAspLeuAsnValAspAsnAsp 1023
3496 ATNAGCGGTTATGCAATAGCGGTTGAGTATTTCCGCCACGCTCA 3545
      ::::
1024 LeuArgProGluAlaGlySerTyrIleAlaAsnLeuAlaAla...As 1039
3546 CAGCGTTTTCGCC...GTACAGGACGAATG..... 3573
      ::::
1039 nThrMetPheThrThrArgLeuHisGluLysGlyAsnThrTyrTyrT 1056
3574 ..GACCGCGTTCGCCGAAGACCGCGCAACGCGGNTTGGACAAGCNCG 3621
      ::::
1056 hTrAspMetValThrGlyGlnGlnLysGlnThrThrMetTrp..... 1069
3622 ATCCGAGACACCAACACACTACGTTGCAAGATTTCGGCGCTACCGCCA 3671
      ::::
1069 ..... 1069
3672 ACAACCGACCTGCGCCAAATGCTATGCAAGAAAAACCTGCGACGCGCG 3721

```

```

1069 ..... 1069
3722 GCCTGGCAGTCCTGTTTCGACACGAGCGAAGAACATTCGACGAC 3771
      ::::: ::::: :::::
1070 ..... MetArgHisGluGlyGlyHisAsnLysTrpArgAsp 1081
3772 GGCATCGGC..... AACTCGGACGCGCTTGCCCGCAG 3803
      ||| ||| ||| ||| ||| |||
1082 GlySerGlyGlnLeuLysThrGlnSerAsnArgTyrValLeuGlnLeuGln 1098
3804 C..... GCGGTTTCGGGCAATACGCGATCGGCGATCGACATCG 3844
      ||| ||| ||| ||| ||| |||
1098 YGlyAspValAlaGlnTrpSerGlnAsnGlySerAspArgTrpHisValG 1115
3845 GCATCGACGACGCGCGCGGTTTACG...ACGGCANTCTTCGACGCGC 3891
      ||| ||| ||| ||| ||| |||
1115 YValMetAlaGlyTyrGlyAsnSerAspSerLysThrIleSerSerArg 1131
3892 ATCGGAGCAAAATCCGCGCGCGCTCGATAC.....GGCATTCGA 3935
      ||| ||| ||| ||| ||| |||
1132 ThrGlyTyrArgAlaLysAlaSerValAsnGlyTyrSerThrGlyLeuTyr 1148
3936 GGCACGATACCGCGCGGTTTCGCGGATCGCATCGAATCGATACATCG 3985
      ||| ||| ||| ||| ||| |||
1148 ValLeuThrTrpTyrAla.....AspAspGluSerArgAsnG 1160
3986 GCGCAGCGCGCTATTTCGCCAAAAGCGGATTCACCGCTACGAAAAGCTC 4035
      ||| ||| ||| ||| ||| |||
1160 YAla.....TyrLeuAspSerTrpArgGlnTyrSerTrpPheAspAsn 1174
4036 AATATCGCACCCCGGCTTCGCTTCACCGNTACGNTACGNTACGNTACG 4074
      ||| ||| ||| ||| ||| |||
1175 ThrValLysGlyAspAspLeuGlnSerGlnSerLysSerLysGlyPhe 1191
4075 ...GCGGCGCATTTAAGCAGATTTATTCATAACCGCGGCAACACATNT 4120
      ||| ||| ||| ||| ||| |||
1191 eThrAlaSerLeuGlnAlaGlyTyrLysHisLysLeuAlaGluPheAsnG 1208
4121 CC.....ATCACNCCTTATTATTCACCTG 4143
      ::::: :::::
1208 LysSerGlnGlyThrArgAsnGlnTrpTyrValGlnProGlnAlaGlnVal 1224
4144 TCCATATACGATGCC..... 4158
      ||| ||| ||| ||| ||| |||
1225 ThrTrpMetGlyValLysAlaAspLysHisArgGlnSerAsnGlyThrIle 1241
4159 .....GCTTCGGGCAAGTCCGAAACGCGCTCAATACCGCNG 4195
      ||| ||| ||| ||| ||| |||
1241 ValHisSerAsnGlyAspGlyAsnValGlnThrArgLeuGlyValLysT 1258
4196 TATTGGCTAG.....GATTTCGGCAAAACCGC..... 4224
      ::::: :::::
1258 ThrTrpLeuLysSerHisLysLysMetAspGlyLysSerArgGluPhe 1274
4225 .....AGTCGGGATGGGCGGCGTAAACCGCGAAATCAAGAGTTT 4262
      ||| ||| ||| ||| ||| |||
1275 GlnProPheValGlnValAsnTrpLeuHisAsnSer.....LysAspPhe 1289
4263 CACGCTGTCNTCCACGCTGCGCGCGCAAGCCGCACTGGAAGCGC 4312
      ||| ||| ||| ||| ||| |||
1289 eSerThrSerMetAspGlyValSerVal.....ThrGlnAspGlyAla 1304
4313 AACACAGCGCGGCGCATCAATTAAGC 4338
      ||| ||| ||| ||| ||| |||
1304 rGAsnIleAlaGlnIleLysThrGly 1312

```

seq_name: p1r2:A43855

seq_documentation_block:

high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
 C:Species: Haemophilus influenzae
 C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

```

C:Accession: A43855
R:Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nonlysozyme
detella pertussis.
A:Reference number: A43855; MUID:92192797
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1536 <BAR>
A:Cross-references: GB:008876; GB:M84616; NID:9475770; PIDN:AAA20527.1; PID:9475771
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:89235, NCBIIP:89239)

alignment_scores:
  Quality: 277.00      Length: 1173
  Ratio: 0.474        Gaps: 58
  Percent Similarity: 49.872      Percent Identity: 20.375

alignment_block:
  US-09-303-518D-651 x A43855 ..

Align seg 1/1 to: A43855 from: 1 to: 1536

127 GGCACACCTTATTCGCGATCACTACCAATACATCGGACTTGGCGCA 176
    |||:::|||||:::||||
349 GlyIleThrTyrLeuGly.....GlyAsn 356
177 AATTAAGCGAAGTTTGCACGCGCGGCGAAGATATTGAGTNTACACA 226
    |||:::|||||:::||||
356 pGluArgGlyGlu.....GlyLysAsnGlyIleGlnLeuAlaLysL 370
227 AAAAAGGAGGTTGTCGCGCAATCAATGACAAACCGCGATGATGAT 276
    |||:::|||||:::||||
370 YS.....ThrSerLeuGlnLysGlySerThrIleAsn 380
277 TTTTCTGTGTGTCGCGTAAGCGCGTGGCGCATTTGGGCGATCAATA 326
    |||:::|||||:::||||
381 ValSerGlyLysGlnLysGlyArgAlaIleValTrpGlyAspIleAl 397
327 TATTGTG.....AGGTCGACATTA 346
    |||:::|||||:::||||
397 AleuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlLysT 414
347 ACGGCGGCTAT.....AACAC 363
    |||:::|||||:::||||
414 hrgIyGlyPheValGlnThrSerGlyHisAspLeuPheIleLysAspAsn 430
364 .....GTGATTTTGGTGGCGAAGCAAGNAATCCGATCAGACCGTTT 407
    |||:::|||||:::||||
431 AlaIleValAspAlaLysGlnTrpLeuLeuAspProAspAsnValSerI 447
408 TTCTTACCAATGTGTGAAGAAATATATTATTAACCGCGAATTCACAC 457
    |||:::|||||:::||||
447 eAsnAlaGlnThrAlaGlyArgSerAsnThrSerGluAspAspGluTyr 464
458 CTTCACAGCGCGAT...TANCATATGCGCGGTTTGCATTAATTTGCACA 504
    |||:::|||||:::||||
464 hrgLysSerGlyAsnSerAlaSerThrProLysArgAsnLysGlnLysThr 480
505 GATCGACAGACCTGTCGAATGACAGATGACATGAGGGGAAATACCTATTC 554
    |||:::|||||:::||||
481 ThrLeuThrAsnThrThrLeuGlnSerIleLeuLysLysGlyThrPheVa 497
555 CGATTAAGAAATAATATCCGAGCGTGCATCGCGCTCAGACACACAC. 603
    |||:::|||||:::||||
497 LAsnIleThrAla...AsnGlnArgIleTyrValAsnSerSerIleAsnL 513
604 .....TATTGGCGTTAT..... 615
513 euserAsnGlySerLeuThrLeuTrpSerGlnGlyArgSerGlyGlyGly 529

```

616GATGATGACAAACAGCGCA 635
530 ValGluIleAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsn 546
636 TTTATCC...TACTCCGGCGCATGTTA.....A 661
546 nleuThrIleTyrSerGlyGlyTyrValAlaSerValHisLysAsnIleSerL 563
662 TTGGC.....GGCAATACACATATGACAGGT..... 687
563 euGIAlaIleGlnLysnIleAsnIleThrAlaLysGlnAspIleAlaPhe 579
688 ...TGGGGAATATATGGCGTATATGATTTGACGGCGCATGTGCGGCATGC 734
580 GluLysGlySerAsnGlnValIleThrGlyGlnGlyThrIleThrSerG 596
735 CAACACATATGGCCCTATGCCATATGACAGTGGCGGAGCGACAGCGTT 784
596 yAsnGlnLysGlyPheArgPheAsnAsnValSerLeuAsnGlyThrGlyS 613
785 CGCCATATGTTATTTATGACAAACAAACATTAATGGCTGCTCAACGGA 834
613 erGlyLeuGlnPheThrThrLysArgThrAsnLysTyrAlaIleThrAsn 629
835 GTTTTACAAACCGCGTACCTTATTCGGCAGGAAACGGTTTCCAGCT 884
630 LysPheGlnGlyThrLeuAsnIleSerGlyLysValAsnIleSerMetVa 646
885 GATACGCAAGATGTGTTCTACGATGACATTTACAGAGCGCATACACT. 933
646 lleuProLysnGlnSerGlyTyrAspLysPheLysGlyArgThrTyrT 663
934ACCGTCNTTTTACACCGCGCATACGACATTTTCCTTTCA 978
663 rPasnLeuThrSerLeuAsnValSerGlnSerGlyLysPheAsnLeuThr 679
979 TCCACACAAACAGCTACG.....GTCACGGTACAGAA..... 1011
680 IleAspSerArgGlySerAspSerAlaGlyThrLeuThrGlnProTyrAs 696
1012 .ACCAACGAAAGGTNTCCAACTCCAAAGCTTAAGTACAGACAGTCCGAC 1060
696 nleuAsnGlyIleSerPheAsnLysAspThrThrPheAsnValGluArg. 712
1061 TGTTCAGCAATCTTTGAATGAACATGATAAGAACAGTTTACGCGCA 1110
713AsnAlaArgValAsnPheAspIleLysAlaProIle..... 724
1111 GGGGTGTTATTCAGTACCGTCCAAAGTTAAACAAGGTGAACCTTTC 1160
725 ...GlyIleAsnLysTyr.....Se 730
1161 TTTTATCGATTACGCAACGCAACATCTATCTATCAACACATCAACC 1210
730 rSerLeuAsnTyrAlaSer.....PheAsnGlyAsnIleSerV 743
1211 AAGCGCGGCGGCTTTGATTATTAAGGTATTTTACGTCCTCCGTA 1260
743 aLserGlyGly.....SerValAspPheThrLeuLeuAlaSer 756
1261 AACACAGAAACGTGCAAGCGCGCGGT..... 1290
757 SerSerAsnVal...GlnThrProGlyValValIleAsnSerLysTyrPh 772
1291 .CATATCAGTGAAGACAGTACCGTTACTTGGAAGTAAGACGC..... 1332
772 eAsnValSerThrGlySerSerLeuArgPheLysThrSerGlySerThrL 789
1333GTGGCAACGAC...CGCTGTCCAAATTCGGCAAA 1365
789 yStrThrGlyPheSerIleGlnLysAspLeuThrLeuAsnAlaThrGlyGly 805
1366 GGACAGCTGACGTTCAAGCCAAAGGGCAAAAC..... 1398
806 AsnIleThrLeuLeuGlnValGlnGlyThrAspLysMetIleGlyLysG 822
1399CAAGCTCGATACAGCTGGCGAGCGAGTACATCTTTGG 1438
822 yIleValAlaLysLysAsnIleThrPheGlnGlyGlyAsnIleThrPhe. 838
1439 ATCAGCAGGCGACGATTAAGCAAAACAGCTTTAGTGAATCGGC 1488
839GlySerArgLysAlaValThrGluIle... 847
1489 TTGNTCAGCGGAGGCGTACGTCACACGATGATCCGATATACATTC 1538
848GluGlyAsnValThrIleAsnAsnAsnAlaAsn..... 858
1539 CCCCAGAACACTATTTTCGCTTTCGCGCGAGCTGTGATTTAAACG 1588
859ValThrLeuIleG 863
1589 GGCATTCGCTTTCGCTCCACCGTATTTCAAAATACC...GATGAAGGCGG 1635
863 LysSerAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspVal 879
1636 ATGATTCNCATCATATATGCCACACAA...ACATCCACGCTTACCAFTTC 1682
880 lleuLeuAsnSerGlyAsnLeuThrAlaGlyLysnIleValAsnIleAl 896
1683 AGCGATGAAGATTTATACACACCGAGTGTGAAGATATC..... 1722
896 agLysAsn.....LeuThrValGlnSerAsnAlaAsnPheLysAlaIleT 911
1723AATACACTTAATTC 1737
911 hrAsnPheThrPheAsnValGlyLysLeuPheAspAsnLysGlyAsnSer 927
1738 AGCAAAAGAAATTCCTACACAGCT...TGTTTGGCGAGAAAGATACGAC 1784
928 AsnIleSerIleAlaLysGlyLysAlaArgPheLysAspIleAspAsnSe 944
1785 CAAACAGCAAGCGCGCTCACCTGTTTACAGCCCGCGCGCAAGAAC 1834
944 rLys.....AsnLeuSerIleThrThrAsnSerSerThrTyrA 958
1835 GCACCCNGCTGTTCCGGCGAGCAAAATTTAAAGCGCAACATCCAGCA 1884
958 rGhrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeuAsnIle 974
1885 ACAACAGCAAACTGTTTTCACGGCGAGACCGACCGCCAGCTTAA 1934
975 ThrAsn.....GluGlySerAspThrGlu.....Me 983
1935 TCATTTAGAGAGCGGCTGTCAAAATGAAGTATCCCAAGAGGAA 1984
983 tGlnIleGlyGlyAspValSerGlnLysGly..... 994
1985 TCGTGTGGACACGACTGATCCGATCCGACGATTTAAACGGAAATTC 2034
995AsnLeuThrIleSerSerLysPylSite 1003
2035 CATATTCAGGCGCGGAGCGGATTTCCCGCAATCTTTGCCAAATGGA 2084
1004 AsnIleThr...LysGlnIleThrIleLysAlaGlyValAspGlyLys 1019
2085 AGCGATTCGATTTGAGCAATCAGCCCAAGACAGATTTTGGTGTGCGCAC 2134
1019 nSerAspSerAlaThrAsnAsnAlaAsnLeu..... 1030
2135 CGCATCAAGCCATACATCTGTACAGTTCGACGTGACGACNGTGTGACA 2184
1031ThrIleLysThrLys..... 1035
2185 AATGTGTGANAANAACATTTACCGACGATAAGATGATTCATTGAC 2234

```

1036 .....GluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheAs 1049
2235 TAAGCAGCATMTNMGCGCANTGTNAGCNCTNMCNATNACNTNNTNMAA 2284
1049 nlysaIagIuIeThrAlaLys..... 1056
2285 ANCTCNCGGCGNTGCNNACTNMAAGCAATCTT...AGTCAAAATGCG 2331
1057 .....AspGlySerAspLeuThrIleGlyAsnThrAsnSerAlaAspGly 1071
2332 GATACAGCTTATACGTACGACCAAGCCCAACCAAAAGCAACCTTAG 2381
1072 ThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSe 1088
2382 CCGTCGGGCAATGCCCAAGCAATTAATCAAGCCCAATTAAGCGCA 2431
1088 fAlaAspGly.....HisLysValThrLeuHisSerL 1099
2432 ACNCATCGGNTGCGGCAATGCTTCATTTAATCAAGCAACAGCGCGCA 2481
1099 ysaValGluThrSerGlySerAsnAsn...AsnThrGluAspSerSerasp 1114
2482 CAAAACGCGAGCTGACGCTTCCGACACAGCTTAAGCAACGTAAGCCA 2531
1115 AsnAsnAlaGlyLeuThrIle.....AspAlaLys...AsnValThr.. 1127
2532 TTCGCGACCAAGCAAGCATGTCCTCGATAGCGATAGGACGATTCCT 2581
1128 .....ValAsnAsnAsnIleThr..SerHisLysAlaVal..... 1138
2582 TTGAAGAAGCCGCTTACCGGACACATCAGCGGACGAGCAAGANACCA 2631
1139 ..SerIleSerAlaThrSerGlyIuIeThrThrLysThrGlyThrThr 1154
2632 TTACACTTAAAA.....GACAGCGAATGAGCGCTCCGCTGAGGACGCA 2675
1155 IleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySerI 1171
2676 ATTAGCAATTTAAACCTTGACACGCGCATTAACCTCAATTCGCGCT 2725
1171 eleuIeGlyIleGlySerSerSerGlySerValThrLeuThrAlaThr.. 1187
2726 ATGCGCAGCATGTCGAGCGCGCAACCGGAGNGTCAAGACAGCGCG 2775
1188 .....GluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrVal 1201
2776 CGCGCGCGTGGCGCGCTTCCTATTCGCTTACACGCGCAACCTTGGT 2825
1202 ThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrI 1218
2826 AGAATCCCGTTTCAACAGCTGACGCTGAACGCGCAATTTGAACNGTCA 2875
1218 eLysGlyThrGlnSerValThrThrSerSerGlnSerGlyIleGly 1235
2876 GAACATTC.....CGCTTATGTCGGAACCTCTC 2904
1235 LysThrIleSerGlyThrValGluValLysAlaThrGlnSerLeuThr 1251
2905 GGCTACCGGAGGACAAATTTGAAGCTGGGAGAAAGTTCGGAAGACTTA 2954
1252 ThrGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnVa 1267
2955 CACCTTGGCGGTC.....AACATTAACGGGCA 2980
1267 ThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrVal 1284
2981 AGAAGCCGTAGCCTCGATCAATGACGGTA..... 3012
1284 snValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIle 1300
3013 .....GTGGAAGGAAAGCAACAACCGCTGCCGAACCTTAATTT 3056
1301 AsnAlaThrGlnGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuTh 1317

```

```

3057 CACCTGCAAAAGCAACAGCTGATGCCGCGCGTGGCTTACCACTCA 3106
1317 rThrGluAlaSerSerHisIleThrSer..AlaLysGlyGlnValAsnLeu 1333
3107 TCCGCAAAAGCGCGAGTCCGCTCGATATTCGGTCAAAAGCAAGG 3156
1334 SerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValTh 1350
3157 CTTTCGCAAAACTCGCAAGGACAGCAAAAGCAAAAGCAAGCAAAAGA 3206
1350 rLeuAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnA 1367
3207 CAACGCGCAAAAGCCTTG 3223
1367 lThrSerGlyThrLeu 1372

```

seq_name: p1r2:G64993

```

seq_documentation_block:
yfal protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: G64993
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64993
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1250 <BLAT>
A:Cross-references: GB:AE00313; GB:U00096; MID:92367132; PIDN:AAC75293.1; PID:q17885
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yfal

```

```

alignment_scores:
Quality: 263.50      Length: 1438
Ratio: 0.422        Gaps: 63
Percent Similarity: 43.394      Percent Identity: 19.402

alignment_block:
US-09-303-518D-651 x G64993  ..
Align seg 1/1 to: G64993 from: 1 to: 1250

```

```

697 AATGGCGTANTTAGTTGAGCGGC.....GATGCGGCCA 731
25 AsnGlyAlaAlaAlaValThrAspSerCysGlnGlyTyrAspValLysAl 41
732 TGCCAAAGCACTATGGCCCTATGCCGATTCGAGGTGCGGAGGAGAC.... 777
41 aSerCysGlnAlaSerArgGlnSerLeuSerGlyIleThrGlnAspTyrS 58
778 .....ACGGTTCGCCAATGTTATTTATGACAAACAAACAATPAA 819
58 erIleAlaAspGlyGlnTrpLeuValPheSerAspMetThrAsnAla 74
820 TGGCTGCTCAACGAGCTTTACAAACCGGCTACCTTATTCG..... 861
75 SerGlyAlaValAlaPheLeuGlnGlnGlyAlaGlnPheSerLeuLeupr 91
862 .GGCAGGAAAGCGTTTCCAGCTGATACCAAGATTGGTCTACGATG 910
91 ogLysGlnThrGlyMetThrLeuPheAlaAsnAsnThrValThrGlyG 108
911 ACATTTACAGAGGC.....GAT 927
108 lutyAsnAsnGlyAlaAlaIlePheAlaLysGlnAsnSerThrLeuAsn 124

```

928	ACACATACGCTGTTTTGAAACCGCGCAGTAACGGACATTTTCC.....	972
125	LeuThrAspAlaIlePheSerGlyAsnValAlaGlyGlyTyrGlyAlaI	141
973TTTACATCCAAACAACACGCTACGGGTAGAGGTAACAGAAACCAACG	1018
141	AlleTyrSerSerGlyThrAsnAspThrGlyAlaVal.....	153
1019	AAAAGTNTCCAAATCCAAAGCTTAAAGTACAGACATCCGACTGTTGAC	1068
154AspLeuArgValThrAsnAlaMetPheArgAsn	164
1069	GAATCTTTGATGAATGAACGTATAAGAACACAGTTTACGGCGAGGGGCTGT	1118
165	AsnIleAlaAsnAspGlyTyrGlyGlyAlaAlaIleTyrThr.....	177
1119	TAATCAGTACCTCTCAAGGTTAAACACACGCTGAACAAACCTTTCTTATCG	1166
178IleAsnAsnAspAlaTyrIleuSerAspValI	188
1169	ATTACGGCAACGGCAAACTCATCTTATTAACAACATCAACCAACGCGCG	1218
188	LePheAspAsnAsnGlnAlaTyrThrSerThrSerTyrSerAspGlyAsp	204
1219	GGCGGTTTGATTTTGAAGCTAT.....	1242
205	GlyGlyAlaAlaIleAspValThrAspAsnAsnSerAspSerLysHisProse	221
1243TTTACGGTC.....TCGGCTTAAACACACGAAACGCTGGCAAG	1279
221	ArgTyrThrIleValAsnAsnThrAlaPheThrAsnThrAlaGln	238
1280	GGCGGGCC.....GTT	1290
238	LysTyrGlyGlyAlaIleTyrThrAsnSerValThrAlaProTyrLeuIle	254
1291	CATATCGAGACAGACGTAACCGCTTACTTGGAAAGTAACAGCGGTGGCAA	1340
255	AspIleSerValAspAspSerTyrTyrGlnAsnGlyGlyValIleValAs	271
1341	CGACCGCTCTCCAAATCGGCAAAAGCGACGCTGCAGCTTCAAGCAAG	1390
271	PgluAsnAsnSerAlaAlaGlyTyrGlyProSerSerAlaAlaG	288
1391	GGGAA.....AACCAAGCTCGATCAGCGTGGCGAC	1422
288	LysIlePheMetTyrIleuGlyLeuSerGlyValThrPheAspIleAlaAsp	304
1423	GGACAGCTATTTTGGATCAGACGAGCAGACGATTAAGGCAAAAAACAACG	1472
305	GlyLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSe	321
1473	CTTTAGTGAATCGGGCTTG.....NTCAGCGCGAGGGGTACGGGCAAC	1516
321	ArgIleAlaGlyThrLeuIleThrLysThrGlySerGlyAspLeuValI	338
1517	TGATTCGCAATAT.....	1530
338	eufsAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGly	354
1530	1530
355	GluValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHi	371
1531CAGTTCAACCCCGCAAAACTCATTTTGGGCTTGGCGGCGAGCGTT	1576
371	scysGlnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerI	387
1577	TGGAT.....TTAAACGGGCACTTGGCTTTCATCCACCTATTCCAAAT	1620
387	LeaspIlnTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln	403
1621	ACCGATGAAGGGCGCATGTTTNCATATATATATGCGACACACATCCAC	1670

```

11721 TCATGACCTTAAATTACACCAAGAAATGGCTACACAGGTGGTGGTT... 17676
435 lelleeluglialaglyglnleuthrillealaglnasnglserrylal 451
1768 .....GGCGAATA.....CATAGACCAAAACGA 17939
452 leualaglyalaglnsermetaleuthrlygaspillevaivalaspa 468
1794 CGGGCGGCTCAACCTGTTTACCAGCCGCCGACAGACCGCACCCGNC 18433
468 pgllyalaval...leuserleuglnllyaspalaalaspleuthrral 484
1844 TGCTTTCCGGCGGACAAAT.....TTAAAGGACATCCAGCAACA 18874
484 euglnasprproglinsertlevalleuansnglygylvalleuasp 500
1888 AACGGCAACTGTTTTCAGCGGACACCGACCGACCGCCTACATCA 19377
501 SerAsp.....PheSerTh 505
1938 TTTAGAGACGGGCTGCATAAATGAGAGTATC.....CCAC 19757
505 rtrpglnserglythrserlythrasnspolyleuclvalserglyser 522
1976 AACGACAATCGTGTGGACAACGACTGGATCCACCGCAGTTTAAACG 20253
522 erglythrvalleuglserrglnasp...valvalaspleualagly 537
2026 GAAATTTCCATTCACGGC..... 20466
538 AspnleuHisllellylaspolyasprglyalTYryalvala 554
2047 .....GGCAGCGGCGTATTTCCGCAATGTTGCCAAAGTG 20839
554 laspalaserasprglnvalserleualasnasnsnserTYrleug 571
2084 AA..... 20855
571 lythrthrclnlealaserglythrleuMetvalSeraspasnserGln 587
2086 ...GGCGATTGCATTTGAGCAATCAGCCGACAGAGTTTGGTGTCG 21322
588 leuglyAspThrHislythrasnarg.....Glnvalillepethrasply 602
2133 ACCGATCAACG.....CATACATCTGTACAGTGTGG 2167
602 sglnnglnserleuMetglnlethrSeraspvalasphrthrSera 619
2168 ACTGACGCGTGCACAAATTTGTGCAANAANCATTAACGACATGA 22117
619 spralalaclyhsnglyasprillegluMetArgalaspclglnval 635
2218 GTGATGCTTCATGTGACTAAGACGACACNTNAGCGGCATGTNAGCTNMC 22657
636 AlavalaspalaglyvalasphrglntrpclyalaleuMet..... 649
2268 CATATNCGTNTTNAANCTCNCGGGCTCNCNCACTNANAGCATC 2317
650 .....AlaspserserGlygln 656
2318 TTATGCAATATGCGATACACGTTTATACAGTCAGCACACCAACCA 23677
656 isglinspnluglyserThr.....LeuThrlys 665
2368 AACGGCAACTTACCTGTGGCGCATCCCAACGACATTTAATTCAGC 24117
:::|||||::: :::: ||| :::

```

666 ThrG_L. AlaGlyThrLeuGlutThrAlaSerGlyTh 678
2418 CACATTAAAGCGCAACNCATCGGNTTCGGGCATCCTTATTATCTAA 2467
 ||||| : : : : : ||||| : : : : :
678 rTrHrInserAlaValrValaIGluGluGlyThrLeuYsgLysPala 695
2468 GCAACAACGCCGCCGCAAAAAGCGCAGTCCGACTTTCCGACACAGCTVAG 2517
 :
695 IAspIleLeuProTyrrAlaSerSerLeuTPvalIGLYaspGLyAlaThr 711
2518 GCAACGTAAAGCATTCGCACTC. AA 2543
 ||||| :
712 PheValrThrgLyAlaAspInAspIleGlnSerIleAspAlaIleSere 728
2544 CGGCATGTCTCCTCGTAGCGGATNAGGCAATATTCATTGTAAGAACC 2593
 :
728 rGIYThrlrIleAspIleSerAspILyThrValIleu. A 740
2594 GCATTACCGGCAACACTCACCGGCAGCAGANACGATTCACCTTAATAA 2643
 ||||| :
740 rIgeurThrgLyGlnAspThrSerValAlaIleuansAla. . . SerLeuPhe 755
2644 GAACGGAATGACGCTGCCGTCAGGCACGGAA. TTAGG 2681
 :
756 AsnGLYaspILyThrLeuValasnaIatThrAspGLyAlalThrLeuThrGI 772
2682 CAATTTAAAC. CTGCAACAG 2701
 ||||| :
772 YgluleAsnrThrAsnLeuGlnThrAspSerLeuthrIyrIleuSerAnv 789
2702 CCACCATTTACACTCAATTCGCCCTATGCCACAGATCTGCAGGCCGCA 2751
 :
789 alThrValasnGLYasnLeuThrAsnrThrSerGLYAlaValSerLeuGln 805
2752 ACCGGAGANGTGCAACACAGCGCGCGCGCTGGCGCGCTCCCTATT 2801
 :
806 AsnGLYAlaIaIeLysp. 811
2802 ATCCGTTACACGCCCAACTTGGTAGAACCCGTTTCAACACGCTGACGG 2851
 ||||| :
812 ThreLeuThr 815
2852 TAAAGCGCAATTTGAACNGTCAAAGAACATTCGCTTATGCGGAACCTC 2901
 ||||| :
815 alAsnGLYaspTYrrThrGLYgLyThrLeuLeuLeuspsrIleuE 831
2902 TTGGGC. TACGGAAGCGCAAAATTGAAGCTGGCGGAAGTTCCGA 2945
 ||| | :
832 AsnGLYaspSpseerValSerAspGlnLeuValMetAsnGLYAsnrThral 848
2946 AGNACTTACACCTTGGCGGTCAACAT. . . ACCGGC. . . AACGAACCCG 2989
 ||||| :
848 agLYAsnrThrThrValaValaIsnsberIleThrgLyIleGlyLupTOT 865
2990 TAAAGCCGATCAATTGACGTAAGTGAAGGAAAGCAACAACCCGCTG 3039
 :
865 hrserThrgLyIleYsrValaValaAspPhalAlaLasprothrGlnPhe 881
3040 TCAGAAACCTTATTTACACCTGCAAAACGAA. . . CACGTGATGCCGG 3086
 :
882 GlAsnsnslagInPheSerIleualagIyserGLYTyrValaIsmeGI 898
3087 CGGTTGGCGTTTACACACTATCCGCAAAAGACGGCGAGTCCGCCCTGCAATA 3136
 ||||| : : | :
898 yAlaTYrasPTyrrThrLeuValJlunspAsnslAspTYrrLeu. . . 913
3137 ATCCGGTCAAAAGAACAGACTTTCCGACAAACTGGCGAAGCAGAACCC 3186
913 913
3187 AAAAACAAGCGGAAAAAGACAAACGCGCAAAAGCTTGACGCGCTATTGCG 3236
913 913

3237	GGCCCGGCGGGATGCCCGCCGAAGAAGACAGAAGGCCTGGCCAAACGGCCG	3286
914ArgSerIInlValThrProPser.	922
3287	GCGNGCGAGCGCGGAAAAATGTCCGATTATTCAGCGCGAGGAAGAA	3356
922	922
3337	AAACGGTGCAAGCGGATTAAGACAGCGCNTTGGGAAACAGCGGAAC	3386
923ProPsoPsProPsProthr	929
3387	GGAACCCGCGCGGNTPACCACGCGCTTCCCCCGCCGCGNCGCCGCGC	3436
929	929
3437	GGGATTTGCCGCAACCGCACGCCCCAACCGCAACTCAACCCCAACCGAC	3486
930ProAsProAsPrProThrProAsPrProThrProAsPr	943
3487	CGGACCTGTATNAGCCGTTATGCGCATATACGCGTTTAGAGTAATTTTCGC	3536
944	ProGIuProThrProAlaTyrlGlnProVaLLeuAsnAlaIysValcIgl	960
3537	CACGCTCAACAGCGTTTTGCGCGTACAGGACGAATTTGACCGCGTGTTC	3586
960	YTYrLeuAsnAsnLeuAraIala.....AsngInAlaPhem	973
3587	CCGAAGACCGCCCGCAAC.....GCNTTTTGACA	3615
973	eMetGelAtgArGaAspRhlIsalaglyIyAspelyelInlrHeusInleu	989
3616	AGCNCACTCCGGAACACCAAACACTACGCTTGCGAAGATTTCCGCGCTA	3665
990	ArgValIlellyIyAspryHrIstYrThrAlaIaglyIyInleua..	1005
3666	CCGCCAACAAACCGACCTGCGCCAATTCGGTATGCAAGAAAACCTCGGCA	3715
1006	...GlnHisgluAspThrSerThValGlInleuSerelyAspLeuPheS	1021
3716	CGCGGGCC.....GTGCGCATCTGTG.....	3735
1021	erelAygrTrpsglyThrAspglyIyUtrPmetleuclyIleValdlyIy	1037
3736	TTTTTGCACAACCGGACCGAACAANTTCGACAGCGCATCGCAACTC	3785
1038	TyrSerAspnsngInglyAspSerIarGerAsnMetThglyThrAraGl	1054
3786	GGCAAGCGCTGCCACAGCGGCCCTTTTCGGCAATTCGGCATTCGGCAGGT	3833
1054	aAspAsnGlnAsnHisIy.....TYrAlaValcIy....	1064
3836	TCGACATTCGGCATTCAGACAGCGCGGGTTTTAGACAGCGCATCTNTCA	3885
1065LeuthSerSerTrpPheGlnHisIyasnGlnIys	1076
3886	GACGCG.....ATCGAGGACA	3904
1077	GlnIyAlaTrpleuAspSerTrpleugInTyralaTrpPheSerAsnIs	1093
3903	AATCGCGCGCGCG.....GTGCGCATTAC.....GGCATTC	3933
1093	pValSerGlnGlnIyAspelyThrasphIstYrHisSerSerelyIleI	1111
3935	AGGCAGATACCGCGCGGTTTC.....GGCGAATTCGCG	3965
1110	IealIsertLeuGlnIyAglYTyrGlnTrpleuProGlyAraGlyAlaV	1126
3970	ATGCACCTGACATCGGCGCAACGCGCTATTTCCGCAAAAACGGATTA	4015
1127	IlelInProGlnAlaGlnValIleTyrgInGlyValcIInGlnAspSprs	1144


```

1029 1aglyleuGlnasnsertPmetAsnPhelIeValAsn..... 1042
1035 997 GGTAAGGTAACAGAAACCAAGAAAGGTTCCATCCA..... 1035
1043 1043 GlysAsnLeuAsnValThrAsnAlaAsnPheserAsnGlnThrProHisG1 1059
1036 1036 .....AAGCTTAAAGTACAGACAGTCCGACTGTTTGACGAACTTTGA 1078
1059 1059 yelrPhesAsnLeuLysAlaAsnAsnIle..... 1068
1079 1079 ATGAAGATGATAAAGAACGATTACGGCGGCGGAGGCTTAAATCAGTAC 1128
1069 1069 ..ThrTrpAspLysSerValSerGlyGlyLysAsnPhelGlyValAsp 1084
1129 1129 CCTCCAGGTTAAACAGCGT.....GAAACCTTCTCTTATATGA 1169
1085 1085 AsnAlaAsnAlaAsnGlyAsnAlaValIleLysAsnValAsnPheserAs 1101
1170 1170 TTACGGCAACGGCAAACTCATCTTA..... 1194
1101 1101 p.....AsnGlyThrLeuIleTyrLysGlyGlyLysAsnSerAlaGlyA 1116
1195 1195 .....TCAACACATCAACCA 1212
1116 1116 snSerLeuThrLeuGlnAsnAsnThrPhesAsnSerTyrAsnIleAsnAla 1132
1213 1213 GGGCGGCGGCTTGTATTT.....GAAGGTGATT 1244
1133 1133 LysAlaGlnAsnLeuIlePhesAsnAsnAsnSerPhesAsnSerGlySery 1149
1245 1245 TACGCTCTGCTGAAACCAACGAACTGGCAGCGCGGCGCTTCATA 1294
1149 1149 rSerPhesAsnAspThrLysAsnValThrPhelGlyThrAsnThrLeu 1166
1295 1295 TCAGTCAACAC.....AGTACGTTACTTGGAAAGTA 1326
1166 1166 leAsnSerAspProPheserArgLeuLysGlySerValSerIleAspAsn 1182
1327 1327 AACGGCTGGCAACGACCGCTGCAAAATCGGCAAGCAGCAGCTGCA 1376
1183 1183 AsnSerIlePhesAsnIleGlyAspLeuThrAspLysThrThrTyrTh 1199
1377 1377 CGTT.....CAAGCAAAAGGGGAAA 1396
1199 1199 rLeuLeuSerGlyAspAsnIleLysTyrAsnAsnGlnAlaLeuAlaAspA 1216
1397 1397 AC.....CAA 1401
1216 1216 snValrPheserLysAsnLeuThrAspLeuIleHisTyrAspGlyGln 1232
1402 1402 GGCTGATGACGCTGGCGAGCGTACAGTATTGATCGACAGCAGAGA 1451
1233 1233 GlyThrLeuLeuArgThrAspAsnAsnThrTyrPhelValGlnPhethrcl 1249
1452 1452 CGATAAAGCAAAAAACAAGCTTATGAA..... 1482
1249 1249 nSerAsnGlyGlnLysPhelValrPhelGlnLutThrPhesAsnProGlySerI 1266
1483 1483 .....ATCGGCTGNTCAGCGGCGAGGCTACGGTGCACACG 1518
1266 1266 leThrTyrLysTyrPhethrIleHisSerSerProPhethrHisThrGlnAla 1282
1519 1519 AATGCGATTAATCAGTTCAACCCGCAAACTTATTGGCTTTCGCGG 1568
1283 1283 AspSerLysAspIleThrAsnGlnValArgLysGlnPhesPhelIer 1299
1569 1569 CGGACGCTTG..... 1578
1299 1299 ogLysThrProValLysValGlyValLysTyrIleAlaProTyrLysA 1316
1579 1579 .....GATTAAACGGGCAT.....TCGCTTTCGTTCCAC 1608

```

```

1316 1316 snGlnAsrLeuIleGlySerSerAlaPhelAlaThrPserLeuAsnPh... 1331
1609 1609 CGTATTCGAATAATACCGATGAAGGGCGCATGATTGNCNATCATATGTCAC 1658
1332 1332 .....GlyAlaThr 1334
1659 1659 AACAAATCCACCGTTACATACAGGGAATGAAGTATTACACAAACGA 1708
1334 1334 rValValGlyThrLeuLeuLeuGlySerAlaGlnGlnLysAlaAsnAla 1351
1709 1709 GTGTAAAGATATTCATAGACTTAATTACAGCAAGAAATGGCTTAAAC 1758
1351 1351 snGlyGlySerIle..... 1355
1759 1759 GGTGGTTGGCGGAGAAATACAGCAACCAAAAGCAAGCGGCGCTCACT 1808
1356 1356 ..ThrPhelGlyAsnAsnLeuLeuTyrLeuHisGlyLysPhesAsnAl 1371
1809 1809 T.....GTTTACCAAGCGGCGGCGAG 1828
1371 1371 aThrAsnIlePhelLeuThrAsnAsnPhesAsnValGlyAsnProAsnAla 1387
1829 1829 AAGACCGCACCCNCGCTTCCGGCGGCAACAAATTTAAACGGCAACATC 1878
1388 1388 .....GlyGlyGlyAlaThrIleAsnPhesAla 1397
1879 1879 ACGCAACA..... 1887
1398 1398 AspLutThrLeuSerAlaAspGlyLeuAsnTyrThrAsnPhelGlnThrVa 1414
1888 1888 .....A 1888
1414 1414 lAlaMetGlyLeuGlnThrSerAlaSerGlnHisSerThrAlaAsnPha 1431
1889 1889 ACGCAACATG..... 1899
1431 1431 snSerLysLeuSerMetGlnIleLysAsnSerAsnPhelArgAspPhethr 1447
1900 1900 .....TTTTACGGCGCAGACGACACCGCGCAGCGCTTACAA 1934
1448 1448 TrpGlyGlyrPhethrPhesAsnSerGlyArgIleThr.....PhelGlnAs 1462
1935 1935 TCATTAGAAAGCGGTGCTCAAAATGAAAGTATCCCAAGAGA.... 1980
1462 1462 nThrThrPheserGlyThrThrAsnIleAsnGlyAlaThrGlySerLys 1479
1981 1981 .....GAAATCGTGGGAGCAACGATCATCNACCG..... 2013
1479 1479 erSerTyrValAsnMetValAlaAsnThrAspLeuIlePhethrAspSer 1495
2014 2014 .....ACGTTAAACGGGAAATTTCCATAT 2039
1496 1496 lleLeuGlyGlyIleArgTyrAspLeuLysAlaAsnIleIlePh 1512
2040 2040 TCAGGCGGCGGAGCGGTG.....ATTCCCGCAATGTTGCCAACTGG 2083
1512 1512 easAsnThrGlnMetValAlaAspValSerLysAsnValAsnGlnSerS 1529
2084 2084 AAGCGATTCGATTTAGC...AATCAGCCCCAAGCAGTTTGTGTGCT 2130
1529 1529 erLeuAsnGlyAsnValThrPhesAsnHisSerArg.....LeuSerVal 1543
2131 2131 GCACCGCATCAAGCCATACATCTGTACAGTTTGGAGCTGGACNGGCTT 2180
1544 1544 LysProAsnAlaIleAlaIleAsnIleGlyLysrPlnThrGlnThrThrLe 1560
2181 2181 GACAAATTTGTGCGAANAANCAATTACCGACAGATTAAGTATGCTTCAT 2230
1560 1560 uGlnAsnAlaSerSerLeuSerPhethrTyrAsnAspSerValAlaAsnPha 1577
2231 2231 TGACTAAGACGACNTNAGCGGCANTGTNAGNCNCCNNTNACGNTNNT 2280
1577 1577 snGlyThrThrAlaPhesAsnGlyValSerTyrLeu..... 1588

```



```

2281 TNAANCTCNCGGCGTGCNCACTMAAGCAATCTTAGTGCAAATG 2330
1589 .....AsnLeuAsnProAsnAl 1594
2331 CGATACACTTATACAGTCAGCCACAGCCACCAAAAGGCAACCTTA 2380
1594 aglnValSerPheAsnGlnAlaAsn....PheAsnAsnAlaAsnValT 1609
2381 GCCTCGTGGGC..... 2391
1609 hrPheTyrGlyIleProLeuPheGlyLysThrProAsnPheGlyAsnSer 1625
2392 .....AATGCCCAAGCAACATTATATCAAGCCACAT 2423
1626 ValArgLeuIleAsnPheLysGlyAspAlaLysPheAsnGlnAlaThrIle 1642
2424 AATC..... 2427
1642 uAsnLeuArgAlaLysAsnIleHisLeuAsnPheGlnGlyAlaSerThrP 1659
2428 .....GGCAACATCGGNTTCGGGCAATGCTCATTT 2460
1659 heGluAsnAsnSerThrMetAsnLeuAlaGluSerSerGlnAlaSerPhe 1675
2461 AATCTA.....AGCAACAAGCCGCACAAAGGAGAGCT 2495
1676 AsnAlaLeuSerValGluGlyLutThrAsnPheAsnLeuAsnGlySerSe 1692
2496 GACGCTTCCGACAAAGCT.....AAGCAAAAGCTAAGCCATT 2533
1692 rLeuLeuSerPheAsnGlyAsnSerValPheAsnAlaProValAsnPheT 1709
2534 CGCACTCAAGCAATGTCCTTACGCGATAGCAAGTATTCATTT 2583
1709 yAlaAsnAsnSerGlnIleSerPheThrHisSerAlaThrPheAsnAla 1725
2584 GAAAACAGCCGCTTACCGCACTCAGCGGACAGCAAGAACAGCAT 2633
1726 AspAlaSerPheAspLeuGlyAsnAsnSerThrLeuAsnPheGlnSerVa 1742
2634 ACACCTTAAAGACAGCAATGACCTCGCCGACAGC..... 2670
1742 lLeuLeuAsnSerAlaLeuAsnLeuGlyAsnGlyLysAsnLeuA 1759
2671 .....ACGGAATTAGCAATTTAACTTACACAGCCACCATTCACATC 2715
1759 lAlaLeuAlaLysGlyAsnPheSerPheGlySerGlnGlyLe...Leu 1774
2716 AATCCGCTTATCGCCACAGATGCTGACGGCAAAACGGCAGNGTTC 2765
1775 AsnLeuSerTyrMetAsnLeuPheGlyLysPheLysAlaSerValT 1791
2766 AGACAGCGCGCGCGCGCTGCGCGCTCCCTA..... 2799
1791 rAspAlaLeuGlnAlaGlnAsnIleAspGlyLeuArgGlyAsnAsnGlyT 1808
2799 ..... 2799
1808 yrgLysIleLeuPheTyrGlyIleGlnIleGlnLysAlaAspTyrSer 1824
2800 .....TTATCGCTTACACCGCACTTCGGTAGA 2838
1835 PheAsnAsnGlyValHisSerTyrPheThrAsnPheLeuAsnThrTh 1841
2829 ATCCGCTTCAACAGCTGACGGTAAAGGCAATTTGAACGTCAA... 2874
1841 rGluThrIleThrGluThrLeuHisAsnAsnArgLeuLysValGlnIle 1858
2875 .....GGACATTCGCTTATGTCGAATC 2901
1858 erGlnAsnGlyAlaSerAsnAsnAlaMetPheAsnLeuAlaProSerLeu 1874

```

```

2902 TTCGCTACCGAGACGACAAATTGAAGCTGCGGAAGTTC..... 2943
1875 TyrAspTyrGlnGlnAsnProTyrAspGluSerLysAsnSerTyrAsnH 1891
2944 .....GAAGACTTACACCTTGGCGGTCACAAATTCGGCA 2980
1891 sThrSerAspLysAlaGlyThrTyrTyrLeuSerSerIleLysGly. 1907
2981 ACGAACCCGTTAGCGCTGCATCAATTGACGAGTACGGAAGGAAGACAC 3030
1908 .....PheGlyLysAsnAsn 1912
3031 AAA...CCGCTGTCGAAACCTT...AATTTCACCTCGCAAAAGCA 3074
1913 GluIleProGlyThrTyrAsnAlaGlnAsnGlnProLeuGlnAlaLeuH 1929
3075 CGTCGATGCCGCGCGTGGCGTTACCAATTCATCCGCAAGGCGAGT 3124
1929 sile..... 1930
3125 TCCGCTGCATTAATCCGCTCAAGACAGACGTTTC.....GACAAA 3168
1931 .....TyrAsnGlnAlaIleSerLysGlnAspLeuAsnMetIleAlaSer 1945
3169 CTCGCAAGCGCAGCAAGCCAAAGCAAGCAAAAGCAAGCGGCAAG 3218
1946 LeuGlyLysGluPheLeuProLysValAlaLysLeuIleAlaSerGlyAl 1962
3219 CTTGACGCGCTGATTCGCGCGCGCGATGCCCGCGCAAAAGACAGAA 3268
1962 aLeuAspAsnLeuAsnLeuAsnSerProAspSerPheGluThrIlePheS 1979
3269 GCCTT 3273
1979 erIle 1980

```

seq_name: p1r2:H98323

seq_documentation_block:

hypothetical protein AGR_L_3085 (imported) - Agrobacterium tumefaciens (strain C58, C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: H98323
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Gold
A.; Liu, F.; Mollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2333-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: H98323
A:status: preliminary
A:molecule type: DNA
A:Residues: 1-1341 <KUR>
A:Cross-References: GB:AE007870; PIDN:AAK90114.1; PID:915160106; GSPDB:GN00170
A:Gene: AGR_L_3085
A:Map position: linear chromosome

alignment_scores:

Quality:	257.00	Length:	1373
Ratio:	0.421	Gaps:	67
Percent Similarity:	44.501	Percent Identity:	19.811

alignment_block:

US-09-303-518D-651 x H98323 ..

Align seg 1/1 to: H98323 from: 1 to: 1341

```

631 GCGGATTTATCTATCCGCGCA...TGGTTAATGCGGCAATACACA 677
151 G1YThrAlaSerPheIleGlyAlaAsnTTPAlaAsnAspGlyAspAlaSe 167
678 TATGACAGGTGGGAAATATATGCGTANTTACTTGGACGGGATGTC 727

```

```

167 rValIglYAsnGlyThrGlySerLeuLeuLeuHisGlyThrThra 184
172 GCGATGCGCAAGCAGTATGGCCATGCGATGCGATGCGCGGAGCG... 774
184 lAserSerGluAsnIle.....TyrValIglYAsnGlySerGlyThr 196
775 .....GACAGCGGTGCGCCAAATGCTTATTTATGCA 803
197 GlySerGlySerLeuLysLeuAspAsnSerSerThrLeuThrValAlaAs 213
804 CAAACA.....ACAAATTAATGCGTCTCAACGAGATTTTACAA 844
213 pThrValIleAlaIglYThrAsnSerThrIleSerSerAla..... 226
845 CCGGATACCGCTTATTCGCGAGAAAGAGCTTCACGCTGATACGCA 894
227 .....AlaGlyGlySerGlyAsnValIglValIleuGlyAla 238
895 GATTGGTTCTAC.....GATGACATTTACAGAGCGGATACACATACG 938
239 SerSerLeuYSerAlaAsnGlyValIleuAlaAsnAspProAspThrVa 255
939 CTATTGGAACCGCGCATACGAGATTTTCTTACATCCACACAA 988
255 IglYThrAlaLeuValSerGlyGlySerSerThrIleSerSerAla 272
989 AC.....GGTACGCGGTACGATACAGAAACCAACGAA 1020
272 spleuValValIglYValIglYValIglYThrLeuThr..... 284
1021 AAGGTTCACATCCAAAGCTTAAGTACAGACAGCTGTTGACGA 1070
285 .....IleThrGlyGlyGlyMetValIleSerGlnThrThrValIleAl 300
1071 ATCTTGAATGAACGTGATTAAGACAGCTTACGCGGCGAGGGGTGTA 1120
300 aAspLeuAsnSerAlaAspIleSerSerValThrValSer..... 313
1121 ATCAGTACCGTCACAGTAAACAGCGTGAACCTTCTTTATTCAT 1170
314 .....ArgLeuGlySerSerLeuGlnSerIleSerSerLeuVal 325
1171 TACGCGACGCGCAACTCATTTATCAACACATCAACGCGCGG 1220
326 ValIglYAsnGly.....GlyAlaAla 332
1221 CGGTTGTATTTGAGAGTATTTACGCTGCTGCTGAACACGAA 1270
332 AlYsLeuValValIglValAlaAlaAlaAlaAlaAlaAlaAlaAla 343
1271 CGTGGCAAGC.....GCGGCGCTTCATATCAGTGAAGACATGAC 1311
344 .....SerGlyGluAlaIleIleIleIleIleIleIleIleIleIle 358
1312 GTTACT.....TGG.....AAGTAAACG 1331
359 ValThrGlyAspGlySerIleSerThrThrGlyAspLeuGlnValIgl 375
1332 CGTGGCAAGCAGCGCTGCTCAAAATCGGCAAGCAGCTGACGTC 1381
375 YAspThrSerAspProGlyIleuAlaGlyAsnGlyThrLeuAsnVal 392
1382 AAGCCAAAGGAA.....AACCAAGCTGCAGCAGC 1413
392 hrAlaIglYSerValAspSerThrValAlaHisLeuGlyValAla 408
1414 GTGGCGAGCGTACAGTATTTGATCAGCAGCAGCAGATTAAGCA 1463
409 GlyAlaThrGlySerAlaIleValAsp.....GlyIleAla 419
1464 AAAACAAGCTTACT.....GAAATCGCTGTCNTCA 1495

```

```

419 sGlySerValThrThrValAspArgAsnSerLeuGluValGlyVal...S 435
1496 GCGCAGGCGTACGTCGCAACTGATGATCCGATTAATCAAGTCAACCCGAC 1545
435 erGlyAlaGlySerLeuAlaAlaThrGlyGlyGlyLeuValAspAlaAla 451
1546 AAATCTATTTGCGCTTCGCGCGGA..... 1572
452 AsnIleIleIleGlyThrAsnThrGlyGlyAsnGlySerValArgAlaSe 468
1573 .....CGTTGGATTTAAAGCGGATCGC 1597
468 rGlyAlaAspSerThrValIleYSerThrArgSerAspLeuAsn...ValGly 484
1598 TTTCGTCACCGCTATTCAAATATCCATGAGAGGCGGATGATGNCAT 1647
484 eutyrglyAsnGlySerMetThrValIglValIglValIleValIleYSer 500
1648 CATAT.....GCCACA.....ACAATCCACCTTACAT 1679
501 ArgAspGlyThrValAlaAlaThrThrGlyGlySerThrSerAlaValThrVa 517
1680 TACAGGGAATGAAGTATTACACAACGAGTGAAGATATCAATAGAC 1729
517 IThrGlyAspGlySer..... 522
1730 TTAATTCAGCAAGAAATTCCTACACGCTGTTGCGGAGAAAGAT 1779
523 .....SerThr..... 524
1780 ACGACCAAAACGAGCGGCGCTACCTGTTTACAGCCGCGCGAGA 1829
525 .....AlaMetThrGlyThrPhePheValIglYThrAlaSerGlyAlaAla 539
1830 AGACCGACCCGCTGCTTCGCGGGA.....ACAATTTAA 1867
539 rGlyAsnValThrValSerAsnGlyGlyAlaIleArgAlaThrGlyAla 556
1868 AC...GGCAGCATCAGCGCAACAAAGGCAACTGTTTTCAGCGGC... 1911
556 hrLeuGlyAspLeuAlaGlyAlaSerGlyThrMetThrIleThrGlyAla 572
1912 .....AGCCGACACCGCCTACAT.....CATTTAAGGAAG 1946
573 GlySerLysValThrAlaThrValAlaAspAsnGlyThrValAsnSerGlySe 589
1947 C.....GGTGTCAAAATGAAGATCCCAAG 1978
589 rValAspValGlyPheGlnGlySerGlySerLeuSerValValAsnGlyG 606
1979 GAGAAATCGTGTGGACACAGCTGATCNACCGCAGCTTTAAAGCGAA 2028
606 lYserLeu.....Asp 609
2029 AATTTCATATTACAGGCGCGGCGTGTATTCGCCAATGTTGCCA 2078
610 AlatyrAsnLeuTyrValGlyAsnAlaLeuGlySerSerGlyAlaValIle 626
2079 AGTGAAGGC...GATTGCAATTTGAGCAATCAGCCCAACGACATTTT 2125
626 uValSerGlyValGlySerHisValSerValAspGlyLeuMetValValG 643
2126 GTGTGCACCGCATCAAGCCATCAATCTGTACACGTTGCGACTGCACN 2175
643 lYAsnAlaGlyAsnGlySerValGluIle.....Thr 653
2176 GGTCTGACAAATGTGTGCAAAAACATTAACGACGATTAAGTATTC 2225
654 GlyGlyAlaSerLeuAlaAlaProThrIle.....LeuIleAla 666
2226 TTCATGTACTAAGACNGACNTNAGCGCANTGNAAGCTNACG 2275
666 a.....ThrGluAlaGlySerThrGlyValLeuSerIleIleValIleGlyS 681

```

```

2276 NTNNNTNAANCTCNCNGGCTGNCNCACTNAAMGCAATCTTAGTGCA 2325
      ::::::::::::::::::::
681 ercIyglInthrAlaArgSerAlaGlyAlaValGluAlaArgAlaIleAla 697
      ::::::::::::::::::::
2326 AATGGCGATACAGCTTATACAGTCACGCAACGCCACCCCAAAAGCGC.. 2373
      ||| ::::::::::::::
698 pheGlyAlaGlyAsnGlySerIleValPheAsnHisSerGluThrGlyTyr 714
      ::::::::::::::
2374 .....AACCTTAGCCTGCTGGCAATGCCCAAGCAACATTTA 2410
      ::::::::::::::
714 rThrLeuSerAlaAspIleSerGlyAlaGlyAlaValAlaGluAlaG 731
      ::::::::::::::
2411 ATCAAGCCACATTTAAACGCAACNCATCGGNTTCGGCAATGCTTCATTT 2460
      ::::::::::::::
731 yValThrThrLeuSerGlyAsnAsnSerTyrSerGlyGlyThrThrIle 747
      ::::::::::::::
2461 AAT.....CTAAGCAACACGCCGCAACAAAGCGCATGTGACGCT 2501
      ::::::::::::::
748 SerAlaGlyMetLeuLysGlyThrAlaLysSerPheGlySerGlyTyr 764
      ::::::::::::::
2502 TTCGCGACAGCTTAAGGCAAAAGTAAAGCATTCGCGACCTCAACAGCGCATG 2551
      ::::::::::::::
764 eValAsnAsnAlaGluLeuValValAspGly.....GlyGlyThrL 778
      ::::::::::::::
2552 TCTCCCTACCCGATAGGCGATATTCATTTTGA..... 2586
      ::::::::::::::
778 euSerAsnAlaIleSerGlyThrGlySerPheGluLysThrGlyAspGly 794
      ::::::::::::::
2587 .....ACAGCGCGCTTACCGGA..... 2604
      ::::::::::::::
795 AsnLeuLeuThrGlyAsnSerThrTyrSerGlyAlaThrAlaValSe 811
      ::::::::::::::
2605 .....CAACTCAGCGCGAGCAAGANACAGATTCACCTTA.. 2640
      ::::::::::::::
811 rAlaGlyLysLeuSerValAsnGlySerLeuAlaSerAlaValSerValG 828
      ::::::::::::::
2641 .....AAA 2643
      ::::::::::::::
828 ySerGlyAlaThrValGlyGlyThrGlyThrIleGlyGlyLeuThrVal 844
      ::::::::::::::
2644 GACAGCGAATGACGCTGCCGTCAGCGACGCAATTAAGCAATTTA..AA 2690
      ::::::::::::::
845 AsnSerGlyGlyThrLeuAlaProGlyAsnSerIleGlyThrLeuThrSe 861
      ::::::::::::::
2691 CCTTGACAGCGCACCATTCACATTCGCGCTAT.....GCCACG 2734
      ::::::::::::::
861 rThrGlyAsnAlaThrPheAlaSerGlySerThrTyrAlaValGluIleA 878
      ::::::::::::::
2735 ATGCTGCAAGC.....GCGCAAAACGCGCAGNGTGCAGAC 2769
      ::::::::::::::
878 sPAlaAspIlySerSerAspArgLeuAlaValThrGlyThrIleThr... 893
      ::::::::::::::
2770 ACGCGCGCGCGCTTCGCGCGCTCCCTATATCCGTTACACCG..... 2814
      ::::::::::::::
894 .....IleAlaAsnAspValSerLeuIleValThrProLeuG 906
      ::::::::::::::
2815 .....CCAACTCGGTAGAAATCCGTTTCAACACGCTGACGATA 2854
      ::::::::::::::
906 yAlaHisSerAlaTyrSerLeuGlyThrArgTyrThrIleLeuThrAla 923
      ::::::::::::::
2855 ACGGCAATTCAGACNGTCAGGACATTCGCTTATGTGCAACTCTTC 2904
      ::::::::::::::
923 hrGlyGlyValThr.....GlyThrPheSerSerValAspIlySerPhe 937
      ::::::::::::::
2905 GCGTACCGCAAGCAAAATTAAGCTGGCGGAAGTCCGAAGACTTA 2954
      ::::::::::::::
938 AlaTyrLeuThrAlaLysValAlaGlnSerGlyAspAspAlaThrThr 954
      ::::::::::::::
2955 C.....ACCTTGCGCGGTCA 2968
      ::::::::::::::
954 rLeuSerPheLeuArgThrSerProAspSerGlyLeuLeuAlaAlaIa 971
      ::::::::::::::

```

```

2969 ACAAATACCGCAACGAACCCGTAAGCCATGATCAATTCAGCGTATGGA 3018
      ::::::::::::::
971 hrSerThrAlaAsnAlaArgAlaAlaAlaAsnAlaValAlaValLeu... 986
      ::::::::::::::
3019 GGGAAAGCAACAAACGCGCTGCCGAACACCTTATTTCAACCCGCAAAA 3068
      ::::::::::::::
987 ...GlyGlnThrSerProLeuTyrGluAlaIleLeuThrLeuGlnGln 1002
      ::::::::::::::
3069 CGAACACGTCATCCGCGCGCTGCTTACCAACTCATCCGCAAGACG 3118
      ::::::::::::::
1002 yGlu.....ThrGlnGlyAlaPheSer...GlnLeuAla.....G 1013
      ::::::::::::::
3119 GCGAGTCCGCTCGATTAATCCGCTCAAAAGCAAGAGCTTTCGACAAA 3168
      ::::::::::::::
1013 yGluIle.....HisProSerLeuAlaMetAlaIleAlaLeuArg 1026
      ::::::::::::::
3169 CTCGGCAAGCGCAAGCCAAAAAACAGCGGAAAAAGACAAACGCGCAAG 3218
      ::::::::::::::
1027 SerGly.....GlnSe 1030
      ::::::::::::::
3219 CCTTGAGCGCGTATTCGCGCGCGCGCGATGCCCGCAAAAGACAGAA 3268
      ::::::::::::::
1030 rArgAspValIleLeuAsnArgLeuArgSerAlaPheGluGlyValAsp. 1046
      ::::::::::::::
3269 GCGTTGCCGAACCGCGCGCGCGCGAGCGGGAATGTCGCATTTATG 3318
      ::::::::::::::
1046 ..... 1046
      ::::::::::::::
3319 CAGCGGAGAGAGAAAAAACGGGTGACAGCGGATTAAGACAGCGCNTT 3368
      ::::::::::::::
1046 ..... 1046
      ::::::::::::::
3369 GGGCAAAACAGCGCAACGGAACCCGCGCGGNTACACCGCCTTCCCC 3418
      ::::::::::::::
1046 ..... 1046
      ::::::::::::::
3419 GCGCCCGCNGCGCCCGGGATTTGCGCGACGCGACGCCCAACGCA 3468
      ::::::::::::::
1047 .....AlaArgProIleLeuPro..... 1052
      ::::::::::::::
3469 CCTACACCCCAACCGCAGCGGACCTGATTAAGCCGTTATGCCAATACGG 3518
      ::::::::::::::
1053 .....ValAlaTyrAlaGluGly 1059
      ::::::::::::::
3519 TTTAGTGAATTTCCGCCACACGCTCAACAGCGTTTCGCCGTACAGAGC 3568
      ::::::::::::::
1059 yGlnAsnProPheAla..... 1064
      ::::::::::::::
3569 AATTGACCGCGGTGTTGCCGAAGACCGCGCAACGCGTTCGACAGC 3618
      ::::::::::::::
1065 .....ValAspAspIlyAlaLeuSerPheTyrPheSer 1075
      ::::::::::::::
3619 NGCATCCGGAACCAACCAACACTACCGTTGCCAAGATTCCGCGCTACCG 3668
      ::::::::::::::
1076 GlyPheArg..... 1078
      ::::::::::::::
3669 CCAACAAACGACACTCGGCCAAATCGGTATGCAAGAAAAACCTGCGACG 3718
      ::::::::::::::
1079 .....SerArg 1081
      ::::::::::::::
3719 GCGCGCTCGGCATCCTGTTTCGACAAACGCGACGCAAAACANCCTTCGAC 3768
      ::::::::::::::
1081 yArgIleAsp.....Ser 1085
      ::::::::::::::
3769 GAGCGCATTCGCAACTGCGACGCGTTGCCACGCGCGCTT...TTCCG 3815
      ::::::::::::::
1086 AspIlyAsnGlySerSerValAspMetAsnGlyGlyGlyLeuPheGln 1102
      ::::::::::::::
3816 GCAATACGCGATCGCGAGGTTTCGACATCGGATCAGACAGCGCGCGGTT 3865
      ::::::::::::::
1102 yLeuAspIly.....GluLeuSerAspSerTyrArgAlaGly 1115
      ::::::::::::::
3866 TTAGCAGCGGCANTCTNTCAGACGCGATCGAGGCAAAATCCGCGCGCG 3915
      ::::::::::::::

```

```

1115 .....||| ||||| |||
1115 aAlaLaLaLaLaLaLaLaLaLaLaLaLaLaLaLeu 1129
3916 GTGGCTCATTACGGCATTGACGATGACGCGCGGTTTCGGGGGATT 3965
1130 AlaAlaSerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1146
3966 CGGCATCGAAGCG.....TACATCGCGCAACGCGCTATTCTGTC 4006
1146 rGnllleGlyProAlaSerLeuAlaGlyLeuAlaAlaAlaAlaAla 1162
4007 AAAAAGCGATTACCGCTACGAAACGTCATATGCGCACCCCGCTTT 4056
1162 lnspllaGluThrArg.....ArgSerIle 1170
4057 GCGTCAACCGNATCCGCGGCGGATTAAGCATATTATTCATAAAC 4106
1171 SerPheSerThrLeuGlnGlnSerLeuSerAlaSerTyrGlySerTh 1187
4107 GCGCAACACATNTCC..... 4122
1187 rSerGlnValAlaPheAlaGluAlaAlaAlaAlaAlaAlaAlaAla 1204
4123 ..ATCACNCTTATTNAGCTGCTCTATACGAGCGCTTCGGGCAAA 4170
1204 lsrleGluProTyrAlaAlaAlaAlaAlaTyr..... 1213
4171 GTCCGAACGCGCTCAAT.....ACCGNGTATGTGCTCA 4205
1214 ValAsnThrArgThrAspGlyPheGlnGluLysGlyAlaAlaAla 1230
4206 GGATTTGCGCAAAACCGC 4224
1230 lSerSerGlySerThrArg 1236

```

seq_name: pir2:C82199

seq_documentation_block:

rxn toxin REXA VC1451 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C.Species: Vibrio cholerae

C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C.Accession: C82199
 R.Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
 Chaidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.
 L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000

A.Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A.Reference number: A82035; MUID:20406833

A.Accession: C82199

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-4558 <HEI>

A.Cross-references: GB:AE004223; GB:AE003852; NID:99655942; PIDN:AN94608.1; GSPDB:GN001

A.Experimental source: serogroup O1; strain N16961; biotype El Tor

A.Gene: VC1451

A.Map position: 1

alignment_scores:
 Quality: 255.00 Length: 1750
 Ratio: 0.332 Gaps: 90
 Percent Similarity: 43.943 Percent Identity: 20.343

alignment_block:

US-09-303-518d-651 x C82199 ..

Align seg 1/1 to: C82199 from: 1 to: 4558

```

37 CGGAAGCCCGAAGACCGCGCATCGCTTCCTCCCTGCTTACTTACG 86
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 ArgAsnAspProGlnThrGlyIleuLysTyr..... 375

```

```

87 CATATGCCCTGCTGGCATTTCTCCCAAGCTTGGCGGACACACTT 136
376 .....TyrSerThrAlaTrp.....T 381
137 ATTTCGGCATCACTAC.....CATACTATCGGCACTTGGCGAA 177
381 TyrLysGluValAsnHisLeuSerAsnLeuAlaAsnGlnAspIleSerAsp 397
178 AATAAGCAAGTTTGCAGTC.....GGGCG.....AAGA 209
398 AsnGlyGlyPheThrAlaValAsnIleAsnGlyAlaTyrThrLeuSerAs 414
210 TATTAGCTTACAAACAAAAAGG.....GAGTTGCTGGCAAT 250
414 PleuLysValGlnHisGlnSerValThrValHisAlaValGlnLys 431
251 CAATGACAAAAGCCCGCATGTTGATTTCTGCGTGTGCGGTAACGC 300
431 erLeuThrGluTyrGluThrPalaThrPala.....AsnGly 443
301 GTGGCGCATGTGGCGATCAATATATGTGACGCGCAGATTAACG 350
444 ..AlaValIleAspAlaLysGluValSerLeuSerPalaLysMetG 459
351 CGGCTAT.....AACACGTTGATTTGGTGGGAG 382
459 YGlyHisAlaIleTyrAlaAspGlyThrLysValAspValLysAlaVal 476
383 GAAGNAATCCGATCAGACCGCTTTCTTCAATTTGGAAGAAAGAAAT 432
476 ySerAsnAlaGlnProAsnThrTyrIleTyrAlaLysVal..... 489
433 AATTATAAGCCTGACAATTACACCTTACACGCGGATTANCATATGCC 482
490 .....LeuGlyProTyr..... 493
483 GCGTTTGCAATAATTGTGCACAGATGCAAGACTGTGAATGACAGTG 532
494 .....ThrLysIleValAlaValGlnLysValAsn 504
533 ACATGAGGCGGATACCATTTCCGATTAAGAAATATCCGACGCTGTC 582
504 sProGluThrGlyAlaLeu.....LysTyrGlnAlaIleSer 516
583 CGCATCGGCTCAGACACCATATTTGGCGTTAT...GATGATCAAAACA 629
517 TrpTyrLysGluGlyAspHisThrAlaAsnIleAlaAsnGlnAspIle 533
630 CGGCGATTATTCCTACTCCGCGCATGTTAATTTGGCGCAATACACATA 679
533 rSerAlaThrGlyTyrAsnPrometGlyLysGlyLysTyrSerLeuSer 550
680 TGACAGGCTTGGGAAATAATGCGCTANTTATGTTGACGCGGAGTGTGCG 729
550 spleuHisTyrSerValAsnAlaValArgSerThrSerGluThrValAla 566
730 CATGCCACGACTATGCCCTATGCGCATTTGCGGCGGACGAGCGACAG 779
567 AspIleGlnGluTyrThrAspGlnThrLeuPheLysProAlaAsnAsp 583
780 CGGTTCCGCAATGTTTATTATGACAAACAAACAAATTAATAGCTGTCA 829
583 rGly.....GluSerSerGlyAspValArgPhe 593
830 ACGGAGTTTACAAACCGCTACCTTATTCGCGACGAGAAACGTTTC 879
593 snGlyAla.....GlyGly 598
880 CAGCTGATACCAAGATTTGTTCTACGANGACATTTACAGAGCGGATAC 929
599 AsnValIleLys.....SerAsnValThrArgGlyAsnVal 610
930 ACAT.....ACCGTCTNTTTGAACCGCGGACGTA 958

```

```
610 HispHeasnglyglyIleAlaasnValIleuHisserSerInp 627
    :||| :||| :||| :|||
959 AGGACATTTTCCCTTACATCCACACACAGCGGTACGGTAAACA 1008
    ||| :||| :||| :|||
627 heglYasnThrGluPhe.....AsnglyglyIleAlaAlaasn 639
    :||| :||| :||| :|||
1009 GAACCAACGAAGAAGTTCACATCCAAAGCTTAAGTACAGACGTCG 1058
    :||| :||| :||| :|||
640 ValIleValIysSerGlyGluGlnGlyAspGlyThrPheArgGlyAlaGl 656
    :||| :||| :||| :|||
1059 ACTGTTTACGCAATCTTGAATGAAGTGAAT.....AAGAACGATTT 1102
    :||| :||| :||| :|||
656 yLeuAlaasnValIleuValHisGlnSerGlyGlnGlyYsmetAspValT 673
    :||| :||| :||| :|||
1103 ACGCGGAGGGGGGTGTTAATCAGTACCGTCCAAAGTTAAACACGGTCAA 1152
    ||| :||| :||| :|||
673 yTAlaGlyGlyAlaValaValasn...ValIleuValArgLeuGlyAspGlyGln 688
    :||| :||| :||| :|||
1153 AACCTTCT...TTATCGATTACGGCAAC..... 1179
    :||| :||| :||| :|||
689 TyrIleuAlaHisIleuLeuAlaTyrGlyAsnIleSerValGlnIysGlySe 705
    :||| :||| :||| :|||
1180 .....GGCAACTCATCTTATCAACACATCAACCA..... 1212
    :||| :||| :||| :|||
705 rGlyAspSerArgValaValaMetLeuGlyGlyTyrFasnThrHisThrGlnI 722
    :||| :||| :||| :|||
1213 ..GGCGCGGG...GGTTGTATTTT...GAAGGATTTTACGGTCCG 1254
    ||| :||| :||| :|||
722 legIySerGlyAsnGlyLeuThrLeuAlaGlyGlyPheAsnValMet 738
    :||| :||| :||| :|||
1255 CQTGA..... 1262
    :||| :||| :||| :|||
729 ThrGlnValGlyLysGlyAspValAlaAlaValLeuAlaGlyAlaAs 755
    :||| :||| :||| :|||
1263 CAACGAACGCTGGCAGCGCGCGG..... 1287
    :||| :||| :||| :|||
755 nValLeuThrIysMetGlyGlnGlyLeuThrSerGlyMetLeuGlyG 772
    :||| :||| :||| :|||
1288 .....GTTCATATCAGTGAAGAC.....AGTACGGTT 1314
    :||| :||| :||| :|||
772 LyAlaAsnValIleThrHisIleSerAsnAspArgIleLeuSerAsnThr 788
    :||| :||| :||| :|||
1315 ACTTGGAAGTAACGGCGGTGCAACAGCCCTGTCCAAATCGSCAA 1364
    ||| :||| :||| :|||
789 ThrIaValAlaLeuGlyGlyAlaasn...IleuThrIysLysGly 804
    :||| :||| :||| :|||
1365 AGGC...ACGCTGCACGTTCAAGCCAAAGGGAACCAAGGCTCGATCA 1411
    ||| :||| :||| :|||
804 sGlyAsnThrLeuAlaValaMetGlyGlyAlaAsnVal...LeuThrH 820
    :||| :||| :||| :|||
1412 GCGTGGCGGCGGTACAGTC..... 1431
    :||| :||| :||| :|||
820 IsValGlyAspGlyThrThrThrGlyAlaMetValGlyGlyAlaAsnIle 836
    :||| :||| :||| :|||
1431 ..... 1431
    :||| :||| :||| :|||
837 LeuThrIysValGlyAsnGlyAspThrThrGlyIleLeuLeuGlyValGl 853
    :||| :||| :||| :|||
1432 ....ATTTGATCAGCAGCAGCAGTAA.....GCCAAAAAACAG 1471
    :||| :||| :||| :|||
853 yAsnValLeuThrHisValGlyAspGlyGlnThrLeuGlyValaMetGlyA 870
    :||| :||| :||| :|||
1472 CCTTAGGAATCGGCTTGNTCAGCGGAGGGGTAGC.....GTG 1512
    ||| :||| :||| :|||
870 IaAlaGlyAsnIlePheThrIysValGlyAspGlyThrSerIleAlaVal 886
    :||| :||| :||| :|||
1513 CAACGTAAATGCGATTAACGTTCAACCCGACAAACTCTATTTCGGCTT 1562
    :||| :||| :||| :|||
887 MetIleGlyAlaGlyAsnIlePheThr.....HisValGlyG 899
    :||| :||| :||| :|||
1563 TCGGGGCGGACGTTTGATTTAAACGGGCAATTCCTTTCACCGGTA 1612
    :||| :||| :||| :|||
899 uGlyAsnAlaTyrAlaLeuMetGlyGlyLeuGlyAsnValPheThrIysV 916
    :||| :||| :||| :|||
1613 TTCAAAATACCAGTGAAGGGCG...ATGATTCNCATATTAATGCCACA 1659
    :||| :||| :||| :|||
916 aGlyAsnGlyAspAlaLeuAlaLeuMetValaIaGlyAlaAsnValPhe 932
    :||| :||| :||| :|||
1660 ACA.....ACATCCACCGCTTACCATTAACAGGAA 1688
    :||| :||| :||| :|||
933 ThrHisIleGlyAspGlyMetSerValAlaLeuMetLeuAlaLysGlyS 949
    :||| :||| :||| :|||
1689 TGAAGTATTTCACAAACCGAGTGTGAAGAT.....ATCAATA 1726
    :||| :||| :||| :|||
949 nValAlaTThrIysValGlyAsnGlyThrThrIleuAlaIaMetValGlyA 966
    :||| :||| :||| :|||
1727 GACTTAATTACAGCAAGAATATGCTACACGCGTTGTTGGCGGAAA 1776
    :||| :||| :||| :|||
966 snValAsnIlePheThrHisIleGlyHisGlySerThrPhe..... 979
    :||| :||| :||| :|||
1777 GATACGACCAAAACGAAGCGGCGCTCAACCTGTTTACACGCGCGCGC 1826
    :||| :||| :||| :|||
980 .....AlaAlaMetIleGlyGlnAlaAsnIleMetThrIysValGlyS 994
    :||| :||| :||| :|||
1827 AGAAGACCGCACCCNCTGCTTCCGGCGGAACAAATTAAACGGCACA 1876
    :||| :||| :||| :|||
994 nAspLeuThrAlaAlaLeuMetValaGlyLysAlaAsnIleMetThrHisV 1011
    :||| :||| :||| :|||
1877 TCACGCAACAAACGGCAACGTTTTCACGCGCGCAGACCGACA..... 1920
    :||| :||| :||| :|||
1011 aGlyAspGlyThrSerLeuGlyLeuPheAlaGlyIleValaValaMet 1027
    :||| :||| :||| :|||
1921 .....CCGCA 1925
    :||| :||| :||| :|||
1028 ThrIysValGlyAsnGlyThrThrIleuAlaIaMetPheGlyLysAlaAs 1044
    :||| :||| :||| :|||
1926 CGCCTACATCATTTAGGAAGCGGTGTCAAAATGGAAGTATCCAC 1975
    :||| :||| :||| :|||
1044 nIleMetThrHisValaGlyAspGlyLeuThrGlyValaLeuAlaLeuGly 1061
    :||| :||| :||| :|||
1976 AAGGAGAAATCGTGGGACACAGCTGATCAACCGCAGTTTAAAGCG 2025
    :||| :||| :||| :|||
1061 IuAlaAsnIleVal.....ThrIysLeuGly 1069
    :||| :||| :||| :|||
2026 GAAATTTCCATATTCAGGCGGCGGCGAGGTGATTTCCCGCAATTTGC 2075
    :||| :||| :||| :|||
1070 AspAspPhe.....MetGlyValaAlaAlaAlaIaLysAlaAsnVala 1084
    :||| :||| :||| :|||
2076 CAAGTGAAGCGGATGN..... 2094
    :||| :||| :||| :|||
1084 IThrHisValaGlyAspAlaTThrThrAlaAlaValaLeuAlaGlyLysGlyA 1101
    :||| :||| :||| :|||
2094 ..... 2094
    :||| :||| :||| :|||
1101 snIleLeuThrIysValaGlyGlyGlyThrThrValaGlyLeuLeuIleSer 1117
    :||| :||| :||| :|||
2095 .....CATTTGAGCAATCAGCGC.....CAAGCATTTTGGTGT 2129
    :||| :||| :||| :|||
1118 AspValaGlyAsnValaMetThrHisValaGlyAspGlyThrThrIleGlyI 1134
    :||| :||| :||| :|||
2130 CGCAGCGCATCAAAAGCCATPACAACTGTACAGCTTCGAC..... 2169
    :||| :||| :||| :|||
1134 eAlaLysGlyLysAlaAsnLeuIlePheThrIysValaGlyAspGlyLeuGlyV 1151
    :||| :||| :||| :|||
2170 .....TGGACNCGTGCACAAATTTGTCTGCAAAANCAATTACC 2208
    :||| :||| :||| :|||
1151 aLysnValIThrTrp...GlyGlnAlaAsnValaPheThrGlnValaGlyAsp 1166
    :||| :||| :||| :|||
2209 GACGATAAGTGTATGCTCA.....TTGACTTAAGC 2240
    :||| :||| :||| :|||
1167 GlyAspArgTyrAsnPheAlaLysGlyLysAlaAsnLeuIleThrIysVala 1183
    :||| :||| :||| :|||
2241 NGACNTNAGCGCGCANTGTNAGCNTNCCNATNAGCNTNNTTNAANCNTCN 2290
    :||| :||| :||| :|||
1183 IglyAspGlyGlnGlnValaIserVal.....ValG 1193
```

2291 CNGGGCTGCGNCCATNMAANGCAGCATCTTAGTGCAGATGGGATACACGT 2340
1193 InGlyGluAlaAsnIleIleThrHisVal...GlyAsnGlyAspAsp... 1207
2341 TATACA.....GTCACCCACAA 2357
1208 TyrThrGlyAlaIleTrpGlyLysAlaAsnValIleThrLysValGlyHisGly 1224
2358 C.....GCCACCCAAACG 2371
1224 yGlnAsnValValLeuAlaLysGlyGluAlaAsnIleValThrGlnValG 1241
2372 GCAC.....CTTACG 2382
1241 LyspGlyAspSerPheAsnAlaLeuTrpSerLysGlyAsnIleValThr 1257
2383 CTGCTGGGCAT.....GCCACGACACATTAAATCA 2414
1258 LysValGlyAspGlyMetGlnValThrAlaAlaLysGlyGlnAlaAsnI 1274
2415 AGCCACATTAAAGCGCAC.....NCATCGGNTCCGGCATG 2452
1274 eTrhTrThrValGlyAsnGlyLeuAsnValThrAlaAlaTrpGlyLys 1291
2453 CTTCATTATAT.....CTAAGCAACAGCGCGCA... 2481
1291 LAsnIleAsnThrLysValGlyAspGlyValSerValAsnValAlaTrp 1307
2481 2481
1308 GlyLysTyrAsnIleAsnThrLysValGlyAspGlyLeuAsnValAlaVal 1324
2482CAAAAGCGCAGTCTGACGCTTCCGACACCCCTAAGG 2518
1324 MetLysGlyLysAlaAsnAlaAsnIleHisValGlyAspGlyLeuAsnI 1341
2519 CAACGTAAGCCATTCCGCATCAAC.....GCGCAT... 2550
1341 LeAsnAlaSerTyrAlaGlnAsnAsnValAlaIleLysValGlyAsnGly 1357
2551GTCGCCCTAGCC.....GATTAAGCG 2570
1358 AspPheLysSerLeuAlaValAlaSerSerAsnThrSerSerAsnLysLe 1374
2571 AGTATTCATTTTGAACACGCCGCTTACGGACACTACGCGCAGCA 2620
1374 uSerAlaLeuPheAspAsnIleLysGlnThrValLeuGlyValGlyLys 1391
2621 AGGACACAGCATTACCTTAAAGACAGGATGAGCGTCGCGCAGCG 2670
1391 eTrGlnAlaIleAsnTyrLeuValGlnGlyAspGluAlaSerSerSerGly 1407
2671 ACGGAATTAGCAATTTAAACCTTGACACAGCCACCATTAACACTCAATTC 2720
1408 ThrHisLysGlyArgGlyAlaIleAlaThrProGluIleThrLysLeuAs 1424
2721 CGCCTATCGCCACGATGCT.....CGAGGCG 2746
1424 pGlyPheGlnMetAspAlaIleLysGluValSerSerAspLeuGlyAsp 1441
2747 CGCAACACCGCAGNNGTCA.....GACACGCCGCGCGCGCTTCCGCG 2790
1441 eTrLeuThrGlySerValThrLysValAspThrProAspLeuAsnLysMet 1457
2791 CGTTCCCATTTATCGTTACCGCCACCACTTCGCTAGATCCGTTTCA 2840
1458 GlnHisAlaLeuAsnVal...AspAspSerSerValGlnAla.....Pr 1471
2841 CAGCGTAGAGGTAAAGCGAA..... 2862
1471 oAsnLeuIleValAsnGlyAspPheGluLeuGlyGlnHisGlyTrpGln 1488


```

406 nglySerGlyAspIleAla.....LysThr.... 414
      :::::|||||
815 ATAAATGGGTGCTCAACGAGTTTACAAACGGCTACCCCTATTCGGC 864
      :::::|||||
415 .....GlyGlyPheValIleThrSerGlyHisIleSerLys 426
      :::::|||||
865 AGGAAACGGTTCAGCTGATACGCAAGATGGTCTTACAT..... 909
      :::::|||||
427 IleGluSerAsnAlaIleValIleLysThrIleLysPheLysAsn 443
      :::::|||||
910 .....GACATTTACAGAGCGCATACATACCG 937
      :::::|||||
443 pAspValThrIleGluAlaGluAspProLeuArgAsnThrGlyIle 460
      :::::|||||
938 TCNTTTTACACGGCGAGTATTTTCCTTACATCCAAAC 987
      :::::|||||
460 snAspGluPheProThrGlyThrGlyAlaSerAspProLysAsn 476
      :::::|||||
988 AACGGTACGGGTACGGTACAGAAACAGAAAGTTCATCCAA 1037
      :::::|||||
477 SerGluLeuLysThr...ThrLeuThrAsnThrIleSerAsnTyrLe 492
      :::::|||||
1038 GCTTAAAGTACAGACAGTCCGACTGTTTACGATCTTTGAATGAACTG 1087
      :::::|||||
492 uLysAsnAlaTyrThrMet.....AsnIleThrAlaSerA 504
      :::::|||||
1088 ATAAAGAACCACTTTACGGCGAGGGGTGTATATCAGTACGCTCAAG 1137
      :::::|||||
504 rGlySerLeuThrValAsnSerSer..... 511
      :::::|||||
1138 TTAAACACGGGTGAAACCTTTCTTTATCGATTACGGCAACGCAACT 1187
      :::::|||||
512 IleAsnIleGlySerAsnSerHisLeuIleHisSerLysGly..... 526
      :::::|||||
1188 CATCTTATCAAAACATCAACCAAGCGCGCGCTTTGTATTGTAAG 1237
      :::::|||||
527 .....GlnArgGlyGlyIleValGlnIleAspG 536
      :::::|||||
1238 GTGATTTTACGCTGCTGCGCTGAAACAGAAACGTCGCAACGCCGGC 1287
      :::::|||||
536 LysPheIleThrSerLysGlyAsnLeuThrIleTyrSerGlyLysTrp 552
      :::::|||||
1288 GTTCATATCATGTAAAGACAGTACGCTTACTGGAAAGTAAACGGCGTGC 1337
      :::::|||||
553 ValAspValHisLysAsnIleThrLeu..... 561
      :::::|||||
1338 AAACGACCGCTGTCCAAATCGGCAAAGCAAGCTGCACGTTCAAGCA 1387
      :::::|||||
561 ..... 561
      :::::|||||
1388 AAGGGGAAACCAAGGCTCGATCAGCTGGCGAGGATCATCTTTTG 1437
      :::::|||||
562 .....AspGlnGlyPheLeuAsnIleThrAlaIleSerValAlaPhe 575
      :::::|||||
1438 GATCAGCAGGACAGCATTAAGCAAAACAGCCTTTGAATGCGG 1487
      :::::|||||
576 ...GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.. 590
      :::::|||||
1488 CTTGNTCAGCGGCGAGGTACGTCACACTGATGCCATATCATGCTCA 1537
      :::::|||||
591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys... 603
      :::::|||||
1538 ACCCGCAAACTCTATTTCGGCTTCGGCGGAGCTTTGGATTAAAC 1587
      :::::|||||
604 .....AspPheArgAlaAsnAsnValSerLeuAsn 613
      :::::|||||
1588 GGGCATTCGCTTTCCTCCACCGTATTCAAATACGATGAAGGGCGCAT 1637
      :::::|||||
614 GlyThrGlyLysGlyLeuAsnIleIleSerSerValAsnAsn..... 627
      :::::|||||
1638 GATTNCATATCATATGCGACACAAACATCCACCGTTACCATTCAGGGA 1687
      :::::|||||

```

```

628 ....LeuThrHisAsn.....LeuSerGlyThrIleAsnIleSerGly 641
      :::::|||||
1688 ATGAAGTATACACACCGAGTGTAG..... 1716
      :::::|||||
641 snIleThrIleAsnGlnThrThrArgLysAsnThrSerTyrTrpGlnThr 657
      :::::|||||
1717 .....AATATCAATAGACTTAAATTACAGCAAAAGAT 1748
      :::::|||||
658 SerHisAspSerHisTrpAsnValSerAlaLeuAsnLeu.....GluThr 672
      :::::|||||
1749 TGCCCTACACGGTTCGTTTGGCGAGAAAGATACGACCAAAACAGCGGC 1798
      :::::|||||
672 rGlyAlaAsnPheThrPheIleLysTyrIleSerSerAsnSerLysGly.. 688
      :::::|||||
1799 GCGTCACCGCTGTTACAGCCCGCGCAGCAAGACCGCACCGCTGCTT 1848
      :::::|||||
689 ..LeuThrThrGlnTyrArgSerSerAla..... 697
      :::::|||||
1849 TCCGCGCAACCAATTTAAACGGCAACATCATCCAAACAAACGCAACT 1898
      :::::|||||
698 .....GlyValAsnPheAsnGly.....ValAsnGlyAsnMe 708
      :::::|||||
1899 GTTTTTC..... 1905
      :::::|||||
708 tSerPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProA 725
      :::::|||||
1906 .....AGCGGACACCGCACCGCACCGCTTC..... 1932
      :::::|||||
725 snGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsn 741
      :::::|||||
1933 .....AATCA 1937
      :::::|||||
742 IleThrAlaThrGlyGlyLysSerValPhePheAspIleTyrAlaAsnH 758
      :::::|||||
1938 TTTAGGAAGCGGTGCTCAAAATGGAAGTATCCCAAGAGGAATCG 1987
      :::::|||||
758 sSerGlyArgGly.....AlaGluLeuLysMetSerGluIleA 771
      :::::|||||
1988 TGTGCGCAACGAGCTGATCACCAGCACGTTTAAAGGGGAAATTC... 2034
      :::::|||||
771 snIleSerAsn.....GlyAlaAsnPheThr 779
      :::::|||||
2035 .....CATATTCAGGCGGCGAGCGCGTGTATTCGCCAATGTTGC 2075
      :::::|||||
780 LeuAsnSerHisValArgGlyAspAlaPhe..... 790
      :::::|||||
2076 CAAGTGGAGCGGATTCGATTCGATTCGATTCAGCAGCCCAAGCAGTTTTC 2125
      :::::|||||
791 ..LysIleAsnLysAspLeuThrIleAsn..... 799
      :::::|||||
2126 GTGTCGACCGCATCAAAAGCCATATCTGTACAGCTTGCGACTGACN 2175
      :::::|||||
800 .....AlaThrAsnSerAsnPhe... 805
      :::::|||||
2176 GGTCTGACAAATGTTGTGANAANAANCATTCAGCAGATTAAGTATGC 2225
      :::::|||||
805 ..... 805
      :::::|||||
2226 TTCATTGACTAAGACNACNTNAGCGGCAATGTNAGCTNCCNATNAG 2275
      :::::|||||
806 ..SerLeuArgGlnThrLysAspAspPheTyrAspGlyTyrAlaArgAsnA 822
      :::::|||||
2276 NTNNTTNAAANCNCNCGGCGNTGCNNCATNTNANGAGCATCTTATGCA 2325
      :::::|||||
822 lAlaIleAsnSerThrTyrAsnIleSerIleLeuGlyLysAsnValThrLeu 838
      :::::|||||
2326 AATGGCATACAGTTATACAGTCAGCCACAAAGCCACCAAAACGCA 2375
      :::::|||||
839 GlyGly.....GlnAsnSerSe 844
      :::::|||||
2376 CTTAGCCTCTGCGCAATGCCAAGCAACATTTAATCAAGCC..... 2418
      :::::|||||
844 rSerSerIleThrGlyAsn.....IleThrIleGluLysAlaAlaAsnV 859
      :::::|||||

```



```

3269 GCGTTGCCGAACCGCGCGGCGNCGAGCGCGGGGAAAATTCGCGATTATG 3318
1064 .....ThrlleGlyAsnSerAspGlyAsnSer 1073
3319 CAGCGCGAGGAGAGAAAAACGGGTGCAGCGCGATTAAAGACAGCGCNTT 3368
1074 GLyAlaGLuAlaLysThrValThrPheAsnAsnValLysAspSerLysIL 1090
3369 GCGGAACACGCGCGAGAGCGAAAC.....CGGCCGANTACCAACG 3409
1090 eSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThrSers 1107
3410 CCGTCCCGCGCGCGCGCGCGCGCGCGGATTGTCGCAACGCGACGCC 3459
1107 eTSerAsnGlyGlyArgGluSerAsnSerAspAsnAspThrGlyLeuThr 1123
3460 CAACGCAACTCAACCCCAACCGCGACCGCGACCTGATNAGC...CGTTA 3506
1124 lIeThrAlaLysAsnValGluValAsnLysAspLleThrSerLeuLysTh 1140
3507 TCCCATACCGCGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTCG 3556
1140 rValAsnLleThrAlaSerGluLysValThrThr..ThrAlaGlySerThr 1156
3557 CCGTACGAGCAGCAATTGACCGCGGTGTCGCGAAGACCGCGCAACGN 3606
1157 lIeAsnAlaThrAsnLlLysAlaSerLleThrThrlLysThrGlyAspIL 1173
3607 GTTTGGACAGAGCGATCCGCGNACCAACACTACCGTTCCG...AAGA 3653
1173 eSerGlyThrLleSerGlyAsnThrValSerValSerAlaThrValAspL 1190
3654 TTTCCGCGCTACCGCCACAAACACCGACCTCGCCAAATGGATTGCAGA 3703
1190 eutThrLysSerGlySerLysLleGluAlaLysSerGlyGluAlaAsn 1206
3704 AAACCTCGGACGCGGCGGTGCGGATCTGTTTGCACMACCGGAC 3753
1207 ValThrSerAlaThrGlyThrLleGlyGly.....ThrLleSerG 1220
3754 GAAMACANCTTCGACGACGCGATCGCA..... 3781
1220 yAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsn 1237
3782 .....ACTCGGACAGGCTGCCACAGCGCGCGTTTCGGGC 3817
1237 lValaGluLleAsnAlaThrGlyGluAlaAlaThrLeuThrAlaThrGly 1253
3818 AAT.....ACGCGATCGGACGAGTTCGACATTCGCGATCAGACGCGCG 3861
1254 AsnThrLeuThrThrGluAlaGlySerSerLleThrSerThrLysGlyG 1270
3862 GGTTTTAGCAGCGCANCTTTCAGACGCGATCGGAGCAAAATCCGCG 3911
1270 nValAspLeuLeuAla.....GlnAsnGlySerLleAlaGlySerLleA 1285
3912 CCGCGTGCATTCAGGCGATTCAGCAGCATACCGCGCGCGTTTCGGCG 3961
1285 snAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrThrValAla 1301
3962 GATTCGCGATTCAGACCGTATGAGCGCAACGCGCTATTTCGTCAAA 4009
1302 GlySerAspLleLysAlaThrSerGlyThrLeuValLleAsnAlaLys 1317

seq_name: pir2:F90073

```

seq_documentation_block: hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kurita, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Kuroda, A.; Matsuda-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, H.

C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: F90073
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2271 <KUR>
 A:Cross-references: GH:BA000018; PID:913702612; PIDN:BA043752.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA2447

alignment_scores:
 Quality: 252.00 Length: 1552
 Ratio: 0.348 Gaps: 51
 Percent Similarity: 46.714 Percent Identity: 17.332

alignment_block:
 US-09-303-518d-651 x F90073 ..

Align seg 1/1 to: F90073 from: 1 to: 2271

```

163 CGCGACTTGGCGAATAAAGCAAGTTGCGAGCGGCGCAAGATAT 212
    |||  ....  |||  |||  |||  |||  |||  |||  |||  |||
    20 ArgLeuTyrLysSerGlyLysAsnTyrValLysSerGlyIleLeuGln 36
213 TCGAGTNTACACAAAGGAGAGTGTGCGCAATCAATGACAAAG 262
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    36 eGIuMeTPhelYsIleMeTcLy.....L 44
263 CCCCGATGATGATTTTCTGTGTGTCGCGTACGGC..... 300
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    44 euProheIleSerHisSerLeuValSerGlnAspAsnGlnSerIleSer 60
301 .....GTGCGCGCATTTGGTGGCGGATCA 323
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    61 LysLysMeTThrGlyTyrGlyLeuYsThrThrAlaValIleGlyAla 77
324 ATATATTGTAGCGTGGCGCATATAC..... 348
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    77 abHeTThrValAsnMeTLeuHisAspGlnGlnAlaPheAlaIleSerAsp 94
349 .....GGCGGCTAT 357
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110
358 AACCAAGTGTATTTGGTGGCGAGAGAAATCCCGATCAGCACCGTT 407
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    111 AsnSerThrThrIleGlnAlaSerThrSerThrAlaAspSer..... 124
408 TTCTTACCAATTTGTCAAAAGAAATATATATTAAGCTGCAATTCACAC 457
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    125 .....ThSerValThrLysAsnSerSerSerValGlnThrSer.... 137
458 CTTACACGCGGATTANCAATATCCGCGTTTGCATTAATTTGTCACAGAT 507
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    138 .....AsnSerAsp.....ThValSer 143
508 GCAGAACTGTGCAATGACGAGTACATGAGGGGAATACCTATTCGA 557
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    144 SerGluLysSerGluLysValThrSerThrThrAsnSerThrSerAsnGln 160
558 TTAAGAAATATATCCGAGCGTGTCCGATCGCTCAGGACACACACTTT 607
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    160 nGlnIuLysLeuThrSerThrSerGluSerThrSerLysAsnThrT 177
608 GCGCTTATGATGATGACAAACGCGGATTAATCTTACTCCGCGGATCG 657
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
    177 hrSerSerSerAspThrLysSerValAlaSerThrSerThrGluGln 193
658 TTAATTGGCGGCAATACACATATGACGGGTTGGGGAATTAATGGCGTANT 707

```

```

194 ProIleAsnThrSerThrAsnGlnSerThrAlaSerAsnThrSerGln 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
708 TAGTTTGCGCGGATGTGGCCCATGCCAAC..... 728
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
210 nSerThrThrProSerSerValAsnLeuAsnLysThrSerThrSerT 227
739 ..GACTATGGCCCTATGCCGATT..... 759
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
227 hrSerThrAlaProValLysLeuArgThrPheSerArgLeuAlaMetSer 243
760 .....GCAGGTCCGCGACGCGCGGT..... 783
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
244 ThrPheAlaSerAlaAlaThrThrThrAlaValThrAlaAsnThrIleThr 260
784 .....TCGCCAATGTTTA 796
260 rValAsnLysAspAsnLeuLysGlnTyrMetThrThrSerGlyAsnAlaT 277
797 TTTATGACAAA..... 807
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
277 hrTyrAspGlnSerThrGlyIleValThrLeuThrGlnAspAlaTyrSer 293
808 .....ACAAACATTAATG 821
294 GlnLysGlyAlaIleThrIleuGlyThrArgIleAspSerAsnLysSer 310
822 GCTGCTACAGAGTATTACAAACCGCTACCCCTATTCGCGAGGAAA 871
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
310 enIhPheSerGlyLysValAsnLeuGlyAsnLysTyrGlnGlyHisGly 327
872 ACGGTTTCAGCTGATACGCAAGATTTGTTTACGATGACATTTACAGA 921
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
327 snGly..... 328
922 GCGGATACATACCGTCTTTTGAACCGCGCATACGAGATTTTTC 971
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
329 GlysPheGlyIleGlyPheAlaPheSerProGlyValLeuGlyIleThrGln 345
972 CTTTACATCCAAC.....AACACGTTACGGGTA 1000
345 yLeuAsnGlyAlaAlaValGlyIleGlyGlyLeuSerAsnAlaPheGly 362
1001 CGGTACAGAAACCAACCAAGAGTNTCCAAATCAAGCTTAAGTACAG 1050
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
362 heLysLeuAspThrTyrHisAsnThrSerLysPro..... 373
1051 ACACTCCGACTGTTTACGATCTTGAATGAATGATTAAGACACAGT 1100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
374 .....AsnSerAlaAlaLysAlaAsnAlaAspProse 384
1101 TTAACGCGGAGGGGTTTAAATCAGTACGTCACAGTTAAACACAGG 1150
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
384 rAsnValAlaGlyGly.....Gly 391
1151 AAAACCTTCTTTATC.....GATTACGCAACGCAACACTATC 1191
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
391 laPheGlyAlaPheValThrThrAspSerGlyAlaAlaThrThrTyr 407
1192 TTATCAACACATCAACACAGCGCGGCGGTTGATTTTGAAGTGA 1241
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
408 ThrSerSerSerThrAlaAspAsnAlaAlaLysLeu..... 419
1242 TTTTACGCTGCGCTGCAAAACCAAGAGTGGACAGCGCGGCGGTTG 1291
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
420 ....AsnValGlnProThrAsnAsn...ThrPheGlnAspPheAspIle 434
1292 ATATCAGTGAAGACACTACCGTT.....ACT 1317
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
434 snTyrAsnGlyAspThrLysValMetThrValLysTyrAlaGlyGlnThr 450
1318 TGAAGAGTAACGCGCTGCAACAGACCGCTGTCCAAATACGCGCAAGG 1367
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

451 TrpHrtrbysnIleSerApTPrIleAlbLySerGlyThrth 465

1368 CACGCTGACGCTTCAACCCAAAGGGGAAACAAAGCGCTGCATCAGC... 1413

465 rAspHserLeuSerMetThrAlaSerThrGlyAlaThrAspLeuG 482

1414GTGGCGAGCGGACGTCGATTTTGGATCAGCAGGCAACGATAA 1455

482 InGlnValGlnPheGlyThrPhe..... 489

1459 GCGAAAAACAAAGCCTTTAGTGAATGGCTTGATCAGCGGAGGGTAC 1508

490GluYrThrGlnSerAlaValThrGlnValArg...Ty 501

1509 GGTGCACGTGAATGGCGATATACGTTCACACCCGACAACTCTATTGC 1558

501 rValAspValThrThrGlyLysAspIleIleProPolysThrThrYserG 518

1559 GCTTTCGCGCGGACGCTTGGATTAAAGGGCATTCGTTCTGTTCAC 1608

518 Ly..... 518

1609 CGATTTCAAATATCCGATGAAGGGCGGATGTCMCNATCATATGCAC 1658

519AsnValAspGln..... 522

1659 AACAACTCCACGCTTACATTTACAGGAATGAAGTATTACAAACGA 1708

523ValValThrIleAspAsnGlnInSerAlaLeuThrAla 536

1709 GTGTAAGATATACATAGACTAATTACAGCAAGAAATATGCTTACAC 1758

536 ysgLyIyrAsnYrThrSerValAspSerSerYrAlaSerThrYrAsn 552

1759 GGTGTGTTTGGCAGAAAGATACGACCAAA.....ACGAAGG 1796

553AspThrAsnYsThrValLysMetThrAsnAl 563

1797 GCGGCTCACCTTGTTCAC.....CAGCCGCGCGCAG 1828

563 agLyGlnSerValThrYrTyPheThrAspAllyAlaProThrVal 580

1829 AAGACGCGACCCGCGTTCGCGGGGACAAATTTAAACGGCAACTC 1878

580 hrValGlyAsnGlnThrIleGluValGlyLysThrMetAsnProVal 596

1879 AGCGAAACAAACGGCAACTGTTTTCACGCGCAACGACGCGCAGC 1928

597 LeuThrThrTrpAsp..... 601

1929 CTACATTCATTATGGAAAGCGGGGTGTCAAAATTTGCAAGTATCCACAAG 1978

602AsnGlyThrGlyThrValThrAsnThrValThrGlyLeuProSerG 617

1979 GAGAAATCGTGTGGGACACGACGTGATCNACCGACGTTTAAAGCGAA 2028

617 Ly.....LeuSerTyAspSerAla 623

2029 AATTTCATATTCAGGGCGGCGGACGCGGTGATTTCCGCCAATGTTGCCAA 2078

624 ThrAsnSerIleLeuIlyThrProThrYrLysIleGlyGlnSerThrValTh 640

2079 AGTGAAGGCGATTCATTAGCAATTCAGCAACCGCCCAACAGCTTTGGTG 2128

640 rValValSerThrAspGlnAlaAsnAsnLysSerThrThrThrPheThrI 657

2129 TC..... 2130

657 IeAsnValValAspThrThrAlaProThrValThrProIleGlyAspGln 673

2131GCACCGCATCAAGCCATCAACATCTACACGTC 2165

674 SerSerGlnValYrSerProIleSerProIleYsIleAlaThrGlnAs 690

2166	GGAC.....TGCAAGCGTCGACAATTGTG	2191
690	pauSsergIyAsnAlaValThrAsnthrValThrglyLeuProSerGIYL	707
2192	TGCANANAACATTAACCGCAGATAAAGTg.....	2220
707	euthrPheaspSerThrAsnAsnthrILeserGIYthrProthrnsIle	723
2221ATTGCTTCATTGACTAAAGCAGCACTNAGCGCANTGTNAG	2261
724	GlyThrSerThrILeserIleValSerThrAspAlasergIY.....	737
2262	NCTNNCCNATNAGCNNTNTTNAANCTCNCNGCGTGNCCNACNTMANG	2311
737	737
2312	GCAATCTTAGTGAATAATGGCGATACACGTTATACAGCACGACACAGCC	2361E
738AsnLySThrThrThrPheylserGIYLVal	748
2362	ACCCAAAAGGC.....AACCTTACCCTCGGGCATGCCAA..	2400E
749	ThraTrgAsnserMetserAspservalSerThrSergIYserThrgIngl	765
2401GCAACATTTAAATCAACCCACATTAACGGCACACNCAT	2437
765	nserGIYserValSerThrSerLysAlaspservInserAlaserThrs	782
2438	CGGNTTCGGGCAATGCTTCAATTATCTACGACACACGCCGACAAAA	2487
782	erThrSergIYserIleValValSerThrSerAlaserThrsIYser	798
2488	GGCAGTCTGACGCTTCCACACACCTTAAGCAACGTAAGCATTCGCG	2537
799	ThrservalserLeuAsperservalSerAlaserIYserLeuSerTh	815
2538	ACTCAACGGCATGTCTCCCTAGCCGATNAGCAGTATTCATTTTGAA	2587
815	rSergIYserAspservalSerSerThrSerLeuValAsns	832
2588	ACGACCGCTTACCGCAACACACGCGGACGAGCAAGAACGATTTACAC	2637
832	erInserValserSerMetserSergIYserValserIYserThrsr	848
2638	TTAAAGACAGCGAATGACGCGCTGCCGTCAGCAGCAATTAGCAATTT	2687
849	LeuSerAspservILeserAsnserAsnserThrgIuIYserGIYser	865
2688	AACCTTGAACACGCCACATTAACACTCAATTCGCCCTATGCCACGATG	2737
865	userrThrSerThrSerAspservLeuArThrsThrSerLeuSerAsps	882
2738	CTGCAGACGCGCAAAAC...GGCAGNNGTGTGAGACACCGCGCGCGCT	2784
882	erLeuSerMetserThrSergIYserLeuSerIYsserGIYserLeuSer	898
2785	TGCGCGCGCTTCCTATATTCGTTACACCGCCAACTTCGGTAGATCCG	2834
899	ThrsrILeserGIYserSerSerThrSerAlaserLeuSerAspserv	915
2835	TTTTCAACACGCTGACGGTAAACGCAAAATTAAGCACTGAAGCAATTC	2884
915	rserAsnAlaIleserThrSerThrSerLeuSergIYserAlaser....	930
2885	GCTTATGTGCGAACTCTTGCGCTACCGAAGCGACAAATTGAACGTGGC	2934
931ThrsrAspservILeserILeser	938
2935	GAAGATTCGGAAGNACTTAACCTTGGCGGTGCMAACATACCGCACAGA	2984
939	AsnserIleAlasnservInserAlaserThrseryIYsserAspservgl	955

```

2985 ACCGGTAAGCCCTGATCAATTGACGGTAGTGGAAGGAAAGCAACAAC 3034
      : : : : : : : : : : : : : : : : : : : : : : : : : :
955 nserThSerIleSerLeuSerThr.....SerAspSerIlys 968
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3035 CGCTG.....TCGGAAGAACCTTATTTACCCCTGCAAAAGCAACAGCTC 3078
      : : : : : : : : : : : : : : : : : : : : : : : : : :
968 ermetSerThSerIleSerLeuSerAspSerThSerThSerIleSer 984
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3079 GATGGCGCGCGCTGGCTTACCACTCCGCAAAAGAGCGGAGTTCCG 3128
      : : : : : : : : : : : : : : : : : : : : : : : : : :
985 ValSerGlySerLeu..... 989
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3129 CCTGCATTAATCCGGTCAAGAAAGAGCTTTCCGACAACCTCGGCAAG 3178
      : : : : : : : : : : : : : : : : : : : : : : : : : :
990 .....SerIleAlaSerGlnSerValSerThSerThSerAspS 1004
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3179 CAGAAGCCCAAAACAGCGCGAAAGCAACGCCCAAGCTTGACCGC 3228
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 ermetSerThSerIleValSerAspSerIleSerThSerIleSer 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3229 CTGATTGGCGCGCGCGATGCGCCGCAAAAGCAAGAGCTTGACCG 3278
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1021 LeuSerAlaSerAspSerIleSerMetSerValSerSerMetSerTh 1037
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3279 ACCGGCGCGCGCGAGCGGGAATGCGCATTAATGAGCGCGAGG 3328
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1037 rSerGlnSerIleSerThSerIleSerLeuSerAspSerIleSer 1052
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3329 AAGAGAAAAACGGGTGACGGCGGATTAAGACAGCGCTTGCGCAAG 3378
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 .....ThSerAspSerAspSer.....LysSer 1060
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3379 CGCGAAGCGGAACCGCGCGGNTACCAACCGCTTCCCGCGCGCGCG 3428
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1061 LeuSerLeuSerThr.SerGlnSerGlySerThSerThSerThSer 1077
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1077 hrSerAlaSerVal.....ArgThSerGlnSerGlnSerThSerGly 1091
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3429 CGCGCGCGCGGATTTCCGCAACCGCACCCCAACCGCAACCTCAACC 3478
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1092 SerMetSerAlaSerGlnSerAspSerMetSerIle..... 1103
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3529 TTTTCCGCGCAGCGCTCAACAGCGCTTTCCGCGTACAGCAATTTGAC 3578
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 .....SerThSerPheSerAspSerThSerAspSerIlys 1116
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3579 CGTGTTCGCGAAGACCGCGCAACGNGTTTGACAAGCNGCATCCGAG 3628
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1116 erAlaSerThrAlaSerSerGlnSerIleSerGlnSerAlaSerThSer 1132
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3629 AACACCAAACTACGCTTCCGAAATTCGCGCGCTACCGCGCAACAAAC 3678
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1133 ThSerGlySerValSerThSerThSerLeuSerThSerAsnSerGln 1149
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3679 GACCTTCGCGCAATCGGTATCGCAAAAACTCGGACAGCGCGCTCG 3728
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1149 unArgThSerThSerValSerAspSerThSerLeuSerThSerGlnS 1166
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3729 CATTCGTGTTT.....CGCACACCGGACGCAAAACACTTCGACGAG 3772
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1166 erAspSerIleSerGlnSerThSerThSerAspSerIleSerGlnAla 1182
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3773 GCATTCGCAACTCGGACAGCTTCCGACGCGCGCTTTCGGGCAATAC 3822
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1183 IleSerAlaSerGlnSerThSerIleSerLeuSerGlnSerAsnSerTh 1199
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3823 GGCATGCGAGGTTCGACATCGCATCAGCAGCGCGGCTTTAGACG 3872
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1199 rSerAspSerGlnSerGlnSerAlaSerIle.....PheLeuSerG 1213
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3873 CGGCANTCTNTCAGACGCGATCGGAGCAAAATCCGCGCGCTGCTGC 3922

```

```

1213 : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1213 lserLeuSerGlnSerThSerThSerThSerThSerIleSerValSerSer 1229
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3923 ATTAGGCAATTCAGCGCATACCGCGCGGTTTCGGCGGATTCGGCATC 3972
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1230 SerThSerIleSerThSerIleSerAspSerThSerIleSerIle 1246
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3973 GAACCGTACATCGCGCGCAACCGCGCTATTTCGTCCAAAAGCGATTACCG 4022
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1246 rThSerThSerThSerAsnSerThSerIleSerAlaSerIle...S 1262
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4023 CTACGAAAGTCATATGCGCCACCGCGGTCTTGCTTCAACCGTATCC 4072
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1262 erThSerThSerIleSer.....GlnSerThSerThr 1273
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4073 GNGCGGCG.....ATTAGCGAGATTATTCATTCAACCGCGCGCAACA 4115
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1274 PheIleSerIleSerValSerThSerIleSerMetSerThSerThSer 1290
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4116 CATTTCCATCACMCCTATTATTNAGCTGTCTATACCGATCCGCTTCGG 4165
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1290 rLeuSerAsnSerThSerLeuSerThSerLeuSerAspSerThSerAla 1307
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4166 GCAAAGTCCGACACCGGTCAATACCGCGNGTATTGGCTCAGAT...TTC 4212
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1307 sPheIleSerAspSerLeuSerThSerMetSerThSerAspSerIle 1323
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4213 GCGCAAAACCGCGAGTCGCAATGGGCGTAAACCGCAATCAAGTTT 4262
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1324 SerThSerIleSerAspSerIleSerThSerThSerIleSerIle 1340
      : : : : : : : : : : : : : : : : : : : : : : : : : :
4263 CACG 4266
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1340 rThr 1341

```

seq_name: p1R2:AF0351

seq_documentation_block:

probable autotransporter protein yapa [imported] - Yersinia pestis (strain C092)

C:Species: Yersinia pestis

C:date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AF0351

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AF0351

A:status: preliminary

A:molecule type: DNA

A:residues: 1-1430 <KUR>

A:cross-references: GB:AL590842; PIDN:CA92137.1; PID:q15980853; GSPDB:GN00175

C:genetics:

A:gene: yapa

alignment_scores:

Quality: 250.50 Length: 1190

Ratio: 0.442 Gaps: 55

Percent Similarity: 47.647 Percent Identity: 20.000

alignment_block:

US-09-303-518D-651 x AF0351 ..

Align seg 1/1 to: AF0351 from: 1 to: 1430

```

955 AGTAAGGACATTTTCTTCTTACATCCAAACACAGCGTACGGTACGCT 1004
      : : : : : : : : : : : : : : : : : : : : : : : : : :
402 AladlagnlupheasprThAlaIyasn.....AlaIadlagnlute 415
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1005 AACGAAACCAACGAAAGGNTCCATCCAAAGCTTAAGTACAGACAG 1054
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

415 uAlaGluThrAlaAlaLysAla.....IleGluAla 426
1055 TCAGACTGTTTGACGAATCTTGAATGAAGATGATTAAGACAGCTTAC 1104
426 IalysIleThrAspLysAlaValGluAspAlaThrAlaAlaTyrLysGlu 442
1105 GCGGCGGGGGTGTATATCAGTACCGTCCAGGTTTAAACAGCGTGAAA 1154
443 AlaAlaAspLysAlaGluGlnThrLysThrAlaLeuGluAlaGluLys 459
1155 CCTTCTTTATCGATTACGCAAGCGCAACTCATCTTATCAACACACA 1204
459 salLys.....GluAspAlaAspLysLeuValValThrAsnThr 472
1205 TCACCAAGCGCGGCGTGTGATTTTGAAGTGATTTTACG..... 1248
473 .....GlyLeuLeuAsnAspAlaAspGlnAlaLeuGlu 483
1249 .....GTCTGCTGAAAAACAAC.....GAACGTGGCAAGCGCGCG 1286
484 GlnLeuValThrAlaGlnAsnAsnAlaGlnProThrLeuAspLeuProAl 500
1287 CCGTATATCAGTGAAGACAGTACCGTACTTGAAAGTAAAGCGCGTGG 1336
500 alLeaspValThrIleAlaProAlaLysThrGlnAspValIleGluGly 517
1337 CAACGACCGCCTGTCCAAATGGCCAAAGCGCGCTGCAGCTTCACAGCC 1386
517 hSerAlaIleAlaThrGlnValAlaGlyGlyThrGlnAsnValAlaLys 533
1387 AAAGGGAAAAACCAAGCTCGATCAGCGTGGCGACGCTGACGATTTT 1436
534 GlyGlyLysAlaIleAspSerValIleThrLysAspGlyIleValAsnLe 550
1437 GGATCAGCAGCGACGATTAAGCAAAAACAACCTTTAGTGAACG 1486
550 uAlaAlaGlyAlaAsnAlaLysGlyThrGlnValThr..... 562
1487 GCTTNTCAGCGGCGGTACGCTGCACTGATCCGATATACGATTC 1536
563 .....LysGlyThr.....LeuAsnAsnAsnGlyLysAl 572
1537 AACCCGCAAACTCTATTTGGCTTTCGGCGACGCTTGGATTAA 1586
573 AspThrAspThrValVal.....SerThrGluGlyLysLeuValLeuTh 587
1587 CGGCAATTCGCTTCCACCGTATTCAAATACCGATGAGGCGCGA 1636
587 rGlyLysSer.....GluThrAlaIleAlaThrSerThrGlyAla 601
1637 TGATTGNCNATCATATATGCCACACA..... 1662
601 ySvalAlaGluGlyGlyValValThrAlaGlyAspHisSerValIleGlu 617
1663 .....ACATCCACCGTTACCTTACAGGGAATGAAGTATTAC 1700
618 LysMetIleSerSerGlyAsnValThrAlaSerGlyAsnAsnThrIleVal 634
1701 ACAACCGAGTGGTAAGATATCATATAGACTTAATTAC..... 1737
634 IArgAspThrThrIleAsnAspGlyLysLeuSerLeuAlaGlyThrAla 651
1738 ..AGCAAGAAATTCCTACACGGTGGTT.....GGCGAG 1773
651 hTrAlaAsnAsnThrThrPheAsnGlyGlyIlePheSerValGluGlyAsp 667
1774 AAAGTACGACCAAAAGCAAC.....GGGCGCTCAACCTTGTTTACA 1817
668 ThrAlaAlaThrLysThrAsnMetThrGlyLysPheAlaValThrGlu 684
1818 GCCCGCCGACAGACGACCCGCTGCTTCCGGCGGAGCAAAATTAA 1866
684 yAsnAlaLysIleGluGluThrValValSerAlaGlyAspValSerLeuA 701
1867 .....AACGGCAACATCACGCAACAACAGCGCAACTGTT 1902
701 IAspLysAlaThrAlaAsnAsnThrThrLeuThrGlyGlyThrPheAla 717
1903 TTCAGCGGCAACGACGACCGCCTACAAATCATTTAGGAAGCGGTG 1952
718 AlaAlaGlyAspThrAlaValSerAlaThrAsnMetThrGlyGlyThrPh 734
1953 GTCAAAATGAGAGTATCCCAAGAGAAATGTGTGGGACACGACT 2002
734 eThr..... 735
2003 GCATCNACCGACGCTTTAAAGCGGAAATTCATATACGGCGGCGAG 2052
736 ..ValAlaGlyAspThrAlaAlaThrLysThrLysMetThrGlyGlyGlu 751
2053 GCGGTGATTTCCCGCAATGTTGCCAAGTGAA..... 2085
2086 ..GGCGATTGNCATTGTAGCAATCACGCCCAAGCGATTTTGTGTGCGAC 2134
767 aserAspPheSerLeuAlaAspLysAlaThrAla..... 778
2135 CGCATCAAGGCATACATCTGTACAGCTTCGACTGACGACGCTGAC 2184
779 .....AsnAsnThrThrLeuThrAspGlyThrPheThr..... 789
2185 AATTGTGCGAANAACATTTACGACGATTAAGTGATTCCTTCATGAC 2234
790 .....ValAlaGlyAsp.....AlaAlaValThr 797
2235 TAAACGACNTNAGCGCAANTGNTAGCTNCCNATNAGTNNNTNAA 2284
797 rAlaThrAsnMetSerGly..... 803
2285 ANCTCNCGGCGNTGCNACCTNAAAGCAATCTTATGCAATGGCGAT 2334
804 .....GlyLysPheAlaValLysGlyLys 811
2335 ACAGTTATACAGTACGACCAACGCCCAAAAGCGCAACCTTGACCT 2384
812 Alalys.....IleLysAspThrGlnLeuSerAlaGlyAsnPheThrLe 826
2385 CGTGGCAATGCCCAAGCAATTTATCAAGCCACATTTAAACGCAACN 2434
826 uAlaGluAsn.....AlaThrAlaAsnAspThrThrLeuAsnGly.... 839
2435 CATCGGNTTCGGGCAATGCTTCATTTAATCTAGCAACAACGCGCA... 2481
840 .....GlyLysPheAspValSerAsnGluAlaThrAla 850
2482 .....CAAAAGCGCACTGACGCTTTCGACACAGCTAA 2516
851 ThrAsnThrThrIleAsnAsnGlyLeuPheThrLeuLysAspGlyAlaH 867
2517 GCGAAGCTAGCCATTCGCACTCAAGCGCAATGTCCTCCAGCGGATA 2566
867 salAsp...SerThrThrValAsnSerGlyThrPheValMetAlaAspG 883
2567 AGCGAGTATTCATTTGAAAAACAGCGGCTTACCGGACACACGCGG 2616
883 In..... 883
2617 AGCAAGANACAGCATTAACCTTAAAGACAGGAGTGCAGCGTGCCTG 2666
884 SerThrAlaAsnGlyIleGlnLeuValAspSerAlaPheThrLeuAlaSe 900
2667 AGCGACGGAATTTAGCAATTTAACTTGACAAAGCCACCATTTACACTCA 2716
900 rGlyAlaLysAlaSerGly.....IleThrLysL 910

```


A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D91018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1250 <NAY>
A:Cross-references: GB:BA000007; PIDN:BA836539.1; PID:913362585; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509552
C:Genetics:
A:Gene: Ecs3116

alignment_scores:

Quality: 250.00 Length: 1412
Ratio: 0.399 Gaps: 66
Percent Similarity: 44.405 Percent Identity: 19.263

alignment_block:

US-09-303-518d-651 x D91018 ..

Align seg 1/1 to: D91018 from: 1 to: 1250

```

646 TCCGGCGCATGGTTATGCGGCAATACATATGACAGGTTGGGAAA 695
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 ThrGlyMetThrLeuPheAlaAsnThrValSerGlyGluTyrAsnAs 111
696 TAATGGCGTA.....N 706
      |||||
111 nGlyGlyAlaIlePheAlaLysGluAsnSerThrLeuAsnLeuThrAspV 128
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
707 TTAGTTTACGGCGCATGTGCGCATGCCACGACTATGCCCTATGCCG 756
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 AlIlePheSerGlyAsnVal.....AlaGlyGlyTyrGlyAlaIle 142
757 ATTGCGGTTGGCGGACGCGACAGCGGTCGCCAATGTTATTATGACAA 806
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 TyrSerGlyThrAsnAspThrGlyAla..... 152
807 AACAAACATAAATGCTGCTCAACGAGTTTACAAACCGGCTACCTT 856
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 ....IleAspLeuArgValThrAsnAlaValPheArgAsnAsnIleAla 168
857 ATTCGGGACGAGGAAAACGTTTCCAGTATACGCAAGAT...TGGTTC 903
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 snAspGlyLysGlyGlyAlaIleTyrThrIleAsnAspIleTyrLeu 184
904 TAGCATGACATTACAGAGCGCATACATACCGCTTTTGAACCGCG 953
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 SerAspAspValPhe.....As 190
954 CAGTACGAGACATTTTCTTACATCCACACACGAGTACGGT.... 999
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
190 nasAsnGlnAlaTyrThrSerTyrSerAspGlyAspGlyAla 207
1000 ....ACGTACAGAAACCAACGAAAGTNTCCATCCAAAGCTTAAA 1044
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
207 laIleAspValThrAspAsnAsnSerIleSHisPro.....Ser 221
1045 GTACAGATCCGACACTGTTGACGAATCTTGAATGAACCTGATTAAGA 1094
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 GlyTyrThrIle..... 225
1095 ACCAGTTACGCGGACGGGGGTGTAATCAGTACCGTCCAGGTTAAACA 1144
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 ..... 228
1145 ACAGTAACACCTTCTTTATGATTACGCAACGCAACATCATCTTA 1194
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
228 sn.....ThrAlaPhe 231
1195 TCAACAACATCAACCAAGCGCGGCGGTTTGTATTGTAAGTGATTT 1244
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 ThrAsnAsnThrAlaGluGlyTyrGlyGlyAlaIleTyrThrAsnSerAl 248

```

```

1245 TACG.....GTCTCGCTGAAACACCAACGACGTCGC 1276
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 aThrAlaProTyrLeuIleAspIleSerValAspSerTyrSer...G 264
1277 AAGCGCGGGGCTTCATATCAGTACAGACACTACCGTTACTTGGAAAGTA 1326
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 lAsnGlyGlyValLeuValAspGlyAsnAsnSerAlaAlaGlyTyrGly 280
1327 AACGCGGTGGCAACGACCGCTGTCCAAATCGCAAGACGACGTCGCA 1376
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 AspGlyPro.....SerSerAlaIleGlyIlePheMetTyr 292
1377 CGTTCAACCCAAAGGCAACCAAGCTCGATACGCTGGCGACGTA 1426
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 rLeu.....GlyLeuSerGluValThrPheAspIleAlaAspGlyL 306
1427 CAGTCATTTTGATCAGCAGCAGACGATTAAGCAAAAAACAGCCTTT 1476
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
306 yThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerIle 322
1477 AGTGAATCGGCTTG.....ATCAGCGGACGAGGTACGTCACATGAA 1520
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 AlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeuAs 339
1521 TGCCGATAT..... 1530
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
339 nAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluV 356
1530 ..... 1530
356 aThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHisCys 372
1531 CAGTTCAACCCCGCAACACTATTTTGGCTTTCGCGGCGAGCTTTGGA 1580
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
373 GlnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerIleAs 388
1581 T.....TTAAC..... 1587
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
388 pLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGlnThrP 405
1588 ..GGCATTCGCTTTCCTCCACCGT....ATTCAAATATACCGATGAA 1629
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 heAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAspAla 421
1630 GGGCGCATGATTCNCATCAATATGCCCAACAACATCCACGCTTACAT 1679
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 GlyGlyAsnValThrValAsnGlnGlySerPheAlaGlyThrIleGluG 438
1680 TACAGGAATGAAGAATGATTTACACACCGACGATGTAAGATATACATAGAC 1729
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
438 yAlaGlyGlnLeuThrIleAlaGln...AsnGlySerTyrVal.....L 452
1730 TTAAATTACAGCAAGAATTCCTTACACGAGTGTGGTGGCGAAGAT 1779
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 euAlaGlyAlaGlnSerMetAlaLeuThrGly.....Asp 463
1780 ACAGCAAAAACGAGCGCGCTCAACCTTGTTTACACCGCGCGCA.. 1827
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 IleValValAspAlaGlyAlaValLeuSerLeuGluGlyAspAlaAlaAs 480
1828 .....GAAGACCGCAACCGCGCTGCTTCGCGGCGGA.. 1857
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
480 pLeuAlaIleAlaLeuGlnAspAspProGlnSerIleValIleAsnGlyGlyM 497
1858 ..ACAAATTTAAACGGCAACATCAGCAAAACCAACGCAACCTGTTTTC 1905
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
497 etLeuAspLeuSerAspPheSerThrTyrGlnSerGlyThrSerTyrLys 513
1906 AGCGGACAGACCGACACCGACGCTTAC..... 1932
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 AspGlyLeuGluValSerGlySerGlyThrValIleGlySerGlnAs 530

```

1933AATCATTTAGACGGCGTGTCAAAA 1960
530 pvalvalaspleuadlaaglyasnaspmethisileglyaspoly 547
1961 TGGAAAGATATC.....CCACAAGAGAAATC 1986
547 ysaaspolyvaltyrvalvalileaspalaclaspolyglinalvalserleu 563
1987 GTGTGGGACAACGACGTGATCMACCGACGCTTTAAACGGGAAATTTCCA 2036
564 Alasnaspasnnglntyrleuglythrthr.....GI 574
2037 TATTCAGGGGGGCGGCGGATTTCCCAATGTTGCCAAGTGGAG 2086
574 nilealaserglythrleuemetvalseraspasn.....SerleuLeug 589
2087 GCGATTTGCATTTGACATATCAGCCGACAGTGTGTTGGTGCACCG 2136
589 lytyrthrhislytraspasnrg.....Glnvalilepethrleuaspolypro 603
2137 CATCAAGC.....CATCAATCTGTACAGCTGGAGCTG 2171
604 GlnGuservValmetGluilethrleuaspasnvalaspthrargserthrth 620
2172 GACNGGTGTGACAAATGTTGTGCAANAANCATTACGACGATTAAGTGA 2221
620 rthrgluhisglyaspilleglumetargalaspolyglinalvalav 637
2222 TTGCTTCATTAAGTAAACN.....GACNTNAGC 2250
637 alaspaalaglyvalaspthrGlntrpGlyalaleuemetalaaspserser 653
2251 GGCATGTGNAAGCTNCCNATNAGCTNNTNAAANCNCGGCGMTC 2300
654 Glycin...HisGlnaspGlyglyserthrleuthrlystrnGlyalagl 669
2301 NNCATNNAAGGCAATCTTAGTGCNAATGCGGATACAGTATATACGTCA 2350
669 ythrleu.....GluLeuthrAlaserclYthrThrGlnserAlaVal. 683
2351 GCCACACGCCACCCAAAAGGCAACCTTAGCTTCGTGGCAATGCCAA 2400
684ArgvalGlnGlnGlnGlnLeuGln.....GlyaspAlaAla 695
2401 GCAACATTTAATCAAGCCACATTA.....AACGGCAACNCATC 2438
696 AspIlephleprotyrAlaserleuThrvalGlyaspGlyAlaThrPh 712
2439 GGNITCGGCAATGCTTCATTTAATCTAGCAACAACGCGCCACAAAACG 2488
712 evalthrGlyAlaspGlnaspIleGlnserIleaspAlaThrserSerG 729
2489 GCAGTCTGACGCTTCGCAACGCTAAGGCAAGGTAAGCATTC... 2535
729 lyThrIleaspIleaseraspGlyThrvalLeuAlaGlyleuThrGlyGlnasp 745
2536GCACTCACGCGCAATGCTCCCTAGCCGTAAGGCGATAT 2576
746 ThrserValAlaLeuasnAlaserleuPhasnGlyaspGlyThrLeuVal 762
2577 CCAT...TTTGAACACGCGCTTTACCGGACACTACGCGGCGACGAAG 2623
762 lasnAlaThraspGlyValThrleuThrGlyGluLeuasnThrAsnLeug 779
2624 AANACGATTAACCTTAATAAAGACGAATGACGCTCGCCGACGAGC 2673
779 luthr.....AspserLeuThrTyrlleuSerAsnValThr 790
2674 GAATTAGCAATTTAAACCTTGACAAAGCCACCATTAACATCAATTCGC 2723
791 ValasnGlyAsnleuThrAsnThrSerGlyAlaValaserleu..... 804
2724 CTATGCCACGATGCTGACAGGCGCAAAACGCGCAGNGTGTACAGACGC 2773
805GlnasnGlyAlaAlaGlyasp..... 811
2774 CGCGCGCGCTTCGCGCGCTTCCTATTATTCGTTACACCGCCAACTTCG 2823
811 811
2824 GTAGATATCCGTTTCAACAGCGTACGGTAAACGCAATTAAGACNGTCA 2873
812ThleuthrValasnGlyaspTyrrthrGlyG 822
2874 AGGAACATTCGCTTTATGTCGAACTCTTCGCGC.....TACGGAACG 2917
822 yglythrleuLeuLeuaspserGlnleuasnGlyaspaspserValserA 839
2918 ACNAATTTGAGCTGGCGGAAAGTTCGGAAGNACTTACACCTTGGCGGTC 2967
839 spGlnleuValmetasnGlyasnThrAlaGlyAsnThrThrValAlaVal 855
2968 AACAAAT...ACCGCG...AACGAACCCGTAAACCTCGATCAATTGACGCT 3011
856 AsnserIlethrGlyIleGlyIuprothrserThrGlyIleValValA 872
3012 AGTGAAGGGAAGACACAAACCGCTGTCGAAACCTTAATTTCACCC 3061
872 lasphealAlaaspProthrGlnpheGlnasnAlaGlnpheSerL 889
3062 TGCAAAACGAA...CACGTGATGCCGCGCGTGGCGTTACCAACTCATC 3108
889 euAlaGlyserGlyTyrrAlaAsnmetGlyAlaTyrraspTyrrthrleuVal 905
3109 CGCAAGACGCGCGAGTTCCGCTGCATTAATCCGTTCAANAACAAAGACT 3158
906 GluaspAsnaspPtyrleu..... 913
3159 TTCCGACAAACTGCGCAAGCAGAACCAAAAACGAGGGAAGAAAGACA 3208
913 913
3209 ACGCGAAAGCCTTGACGCGCTGATTCGCGCGCGCGCGATGCCCGGAA 3258
913 913
3259 AAGACAGAAAGCGTTGCCGACCGCGCGCGCGCGGAGGGAATATCT 3308
914 ArgserGlnGlyValThrProProser..... 922
3309 CGGCATTATGACGCGGAGAGAGAAAGAAACGCGTGCAGCGGATTAAG 3358
923ProProaspProA 927
3359 ACAGCGCNTTGGGAAACAGCGGGAAGCGGAACCGCGCGCNTATCCACC 3408
927 spProthr..... 929
3409 GCCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGGAGCC 3458
930ProaspProaspPr 934
3459 CCAACCGCAACCTCAACCCCAACCGGACGCGACCTGATNAGCGCTTATG 3508
934 othrGlnaspProaspProthrProaspProGlnProthrProAlaTyrg 951
3509 CCAATAGCGGTTGAGTGAATTTTCCGCAACGCTCAACAGCGTTTCGCC 3558
951 lnpValAlaValasnAlaValGlyGlytyrleuasnAlaValArgAla 967
3559 GTACAGACGAATTTGACCGCGCGCTTTGCCGAGAGACCGCGCGCAAC... 3603
968 Ala.....AsnGlnAlaPheMetGlnuArgTrgaspHisAl 960
3604GCNGTTTGACAAACGNCATCCGGAACCAAC 3637


```

980 aglyglaspiglyglnthrleuasnlleuargvallelyglasptryh 997
3638 ACTACCGTTCGCAAGATTTCCGGCTACCGCCACCAACAGACCTGGCC 3687
997 ISTYThrAlaAlaIleGlnleuAla.....GlnHisGlnAspThrSer 1011
3668 CAATCGGTATGCAAAAAACCTCGGCGGGCGC..... 3723
1012 ThrValGlnleuSerGlyAspLeuPheSerGlyArgTrpGlyThrAspGly 1028
3724 .....GTGCGCATCTGTTTTCGACAAACCGGACCCAAACAACT 3763
1028 ygluTrpMetLeuGlyIleVal.....GlyGly 1038
3764 TCGACGAGCGCATCGCAACTCGGACGCGTTGGCCGCGCGCTTTTC 3813
1038 yTserAspAsnGlnIleAspSerArgSerMetThrGlyThrArgAla 1054
3814 GGGCAATACGGCATCGGACGTTTCGACATCGGCATACAGCGGCGCGG 3863
1055 AspAsnGlnAsnHisGly...TyrAlaValGlyLeu...ThrSerSerTr 1069
3864 TTTTACGACGCGCATCTTTCACAGCGCATCGGA..... 3897
1069 pHeGlnHisGlyLysGlnLysGlnGlyAlaTrpLeuAspAsnTrpLeuG 1086
3898 .....GGCAAAATCGCGCGCGCGCTGGTG 3921
1086 lntYrAlaTrpPheSerAsnAspValSerGlnHisGlnAspGlyValAsp 1102
3922 CATTAC.....GGCATTCAGCAGCATACCGCGCGCGTTTC..... 3957
1103 HisTyrHisSerSerGlyIleIleAlaSerLeuGlnAlaGlyTyrGlnTr 1119
3958 .....GGCGATTCGGCATCGAACCTGACATCGGCGCACCGCGCT 3997
1119 pLeuProGlyArgGlyValIleGlnIleGlnIleGlnIleValIleTyrG 1136
3998 ATTTCGTCCAAAAGCGATTCACGCTACGAAAC..... 4032
1136 lngIyAlaGlnIleAspAsnPhenThrAlaAlaAsnArgAlaArgValSer 1152
4033 .....GTCAATATCGCACCCCGCGCTTGGCTTCACCCGNTA 4070
1153 GlnSerGlnGlyAspAspIleGlnThr..... 1161
4071 CCGNCGGGCATTAAGCAGATTAATTCATCAACCGCGCACACATNT 4120
1162 ArgLeuGlyLeuHisSerGlyTrp....ArgThrAlaValHis.... 1174
4121 CCATACNCCATTATTNAGCCTGTCC...TATACGATGCGCGCTTCGGGC 4167
1175 ..ValIleProThrLeuAspLeuAsnTrpTyrHisAspProHisSerThr 1190
4168 AAAGTCCGACACGCGTCATACCGCNGTATTTGGCTACAGATTTGGCA 4217
1191 GlnIleGlnGluAspAlaSerThr.....IleSerAspAspAlaValIly 1205
4218 AACCCGC...AGTGGGGAATGGGGGTAAACCCGCAATC...AAAGTT 4261
1205 sGlnArgGlyGlnIleLysValGlyAlaThrGlnHisIleSerGlnArgV 1222
4262 TCACGCTGTCNNTCCAGCTGCGCGCAAGGACCGCAACTGGAAGG 4311
1222 alSerLeuArgGlySerValAlaTrpGlnLysGlySerAspAspPheAla 1238
4312 CAACACAGCGCGGCATCAATTAAGCATACCGCTGG 4347
1239 GlnThrAlaGlyPheLeuSerMetThrValIlyStrp 1250
seq_name: p1r2:F85862
seq_documentation_block:

```

```

hypothetical protein yfal [Imported] - Escherichia coli (strain 0157:H7, substrain E
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85862
R:Perna, N.T., Plunkett III, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., May
Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dinalanta, E., Potamousis, K., Apod
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85862
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1250 <STO>
A:Cross-references: GB:AE005174; NID:g12516568; PIDN:AA657362.1; GSPDB:GN00145; UMG
C:Genetics:
A:Gene: yfal

```

alignment_scores:

```

Quality: 250.00 Length: 1412
Ratio: 0.399 Gaps: 66
Percent Similarity: 44.405 Percent Identity: 19.263

```

alignment_block:

```

US-09-303-518D-651 x F85862 ..

```

```

Align seg 1/1 to: F85862 from: 1 to: 1250

```

```

646 TCCGCGCGCATGTTAATTCGCGGCATATACATATGCAAGGCTTGGGAAA 695
95 ThrGlyMetThrLeuPheAlaAsnAsnThrValSerGlyLysArgAsnAs 111
696 TAATGGCGTA.....N 706
111 nGlyIyAlaIlePheAlaLysGlnAsnSerThrLeuAsnLeuThrAspV 128
707 TTAGTTGACGCGGATGTCGCCATGCAACGATATGACCTATGCCG 756
128 alIlePheSerGlyAsnVal.....AlaGlyGlyTyrGlyAlaIle 142
757 ATTGCAGGTGGCGGACGACGCGGTTCGCCAATGTTATTATGCAAA 806
143 TySerSerGlyThrAsnAspThrGlyAla..... 152
807 AACCAACAATTAATGCTGCTCAACGAGATTTCACAACCGGTAACCTT 856
153 ....IleAspLeuArgValThrAsnAlaValPheArgAsnAsnIleAla 168
857 ATTCCGCGCAGGAAACGATTTCACAGCTGATACGCAAAAGT...TGGTTC 903
168 snAspGlyLysGlyAlaIleTyrThrIleAsnAsnAspIleTyrLeu 184
904 TAGCATGACATTTCACAGCGGATACACATACGCTCTNTTTTGACCCG 953
185 SerAspAspValPhe.....As 190
954 CAGTAAAGCAATTTTCCTTACATCCAAACAGGTAAGGCT... 999
190 nAsnAsnGlnAlaTyrThrSerThrSerTyrSerAspGlyAspGlyAla 207
1000 .....ACGTTAACAGAAACCAACGAAAGGTTTCCATTCACAAAGCTTAA 1044
207 laIleAspValThrAspAsnAsnSerSerIyShSpro.....Ser 221
1045 GTACAGACAGTCGACGTGTTGACGAATCTTTGAATGAACCTGATAAGA 1094
222 GlyTyrThrIle..... 225
1095 ACCAGTTTACCGCGCAGGGGTGTTAATCAGTACCGTCAAGGTTAACA 1144
226 .....IleAsnA 228

```

```

1145 ACGGTAAACCTTTCTTTATCGATTACGGCAACGCAACTCATCTTA 1194
1195 TCACACACATCAACCAAGCGCGGGTGTATTTGAAGCGATTT 1244
232 ThrAsnThrAlaGluGlyTyrGlyAlaIleTyrThrAsnSerAl 248
1245 TACG.....GTCTGCCTGAAAAACAAGAAAGCTGC 1276
248 arThrAlaProTyrLeuIleAspIleSerAlaAspSerTyrSer...G 264
1277 AAGCGCGCGCTTCATCATCTAGTAAGACAGTACCTTACTTGGAAAGTA 1326
264 IAsnGlyAlaLeuValAspGluAsnAsnSerAlaIleGlyTyrGly 280
1327 AAGCGCGTGGCAACGACCGCGCTGTCCAAATCGCAAAAGCAACCTGCA 1376
281 AspGlyPro.....SerSerAlaIleGlyLysPheMetTy 292
1377 CGTTCAAGCCAAAGGGAACCAAGCTCGATCAGCGTGGCGACGTA 1426
292 rLeu.....GlyLeuSerGluValThrPheAspIleAlaAspGlyL 306
1427 CAGTCATTTGGATCAGCAGCAGCAGATAAAGCAAAAACACGCTTT 1476
306 yThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerIle 322
1477 AGTGAATACGGCTG.....NTCAGCGCGCAGGGTACGGTCAACTGAA 1520
323 AlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeuAs 339
1521 TGCCGATTAAT..... 1530
339 nAlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluV 356
1530 ..... 1530
356 aThrLeuGlyArGSerAsnSerLeuMetAsnValGlyAspThrHisCys 372
1531 CAGTTCACCGCGCACAACTCATTTGCGCTTTCGCGGCGCGCTTGA 1580
373 GlnAspAspProGlnAspCysTyr...GlyLeuThrIleGlySerIleAs 388
1581 T.....TTAAAC..... 1587
388 pLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGlnThrP 405
1588 ..GGGCATTTCGCTTCGCTTCACCGT.....ATTCAAAATACCGATGAA 1629
405 heAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAspAla 421
1630 GGGGCGATGATGNCNATCATTAATGCCAACAAACATCCACCGTTACAT 1679
422 GlyGlyAsnValThrValAsnGlnGlySerPheAlaGlyThrIleGlnG 438
1680 TACAGGAGATGAAGATTATACACACCGAGTGTAGAATATCATATAGAC 1729
438 yAlaGlyGlnLeuThrIleAlaGln..AsnGlySerTyrVal.....L 452
1730 TTAATTAACGCAAGAAATTCGCTTACACAGGTGGTTGGCGAGAGAGAT 1779
452 euAlaGlyAlaGlnSerMetAlaLeuThrGly.....Asp 463
1780 ACGACCAAAAGCAAGCGCGCTCAACTGTTTACACGCCCGCGCGCA.. 1827
464 IleValValAspAlaGlyAlaValLeuSerLeuGluGlyAspAlaAlaAs 480
1828 .....GAAGACCGCAGCCNCGTGTCTCCGCGCA.. 1857
480 pLeuAlaIleLeuGlnAspAspProGlnSerIleValLeuAsnGlyGlyM 497
1858 ..ACAAATTTAAACGCGCAACATCACGCAAAACGCGCAAACTGTTTTC 1905

```

```

497 etLeuAspLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLys 513
1906 AGCGGACAGCCGACACCGCACGCTAC..... 1932
514 AspGlyLeuGluValSerGlySerSerGlyThrValIleGlySerGlnAs 530
1933 .....AATCATTTAGCAAGCGGGTGGTCANAAA 1960
530 pValValAspLeuAlaGlyGlyAsnAspMetHisIleGlyAlaAspGlyL 547
1961 TGAAGGTATC.....CCACAAGAGAAATC 1986
547 yAspGlyValTyrValValIleAspAlaGlyAspGlyGlnValSerLeu 563
1987 GTGTGGCAACAGCACTGATCNCGCCGACGTTTAAAGCGGAAATTTCCA 2036
564 AlaAsnAspAsnGlnTyrLeuGlyTyrThr.....G 574
2037 TATTCAGGCGCGCGCGGATGATTTCCGCAATGTTGCCAAAGTGAAG 2086
574 nIleAlaSerGlyThrLeuMetValSerAspAsn.....SerGlnLeuG 589
2087 GCGATTNCATTTGAGCAATACAGCCCAAGCAGTTTGTGTGCGACCG 2136
589 LysTyrThrHisTyrAsnArg.....GlnValIlePheThrAspLysPro 603
2137 CATCAAGC.....CATCAATCTGTACAGTTCGGGAC 2171
604 GlnGluSerValMetGluIleThrAlaAsnValAspThrArgSerThrTh 620
2172 GACNGGTGTGACAAATTTGTGCAANAAACATTTACGACGATTAAGTGA 2221
620 rThrGlnHisGlyArgAspIleGluMetArgAlaAspGlyGluValAlaV 637
2222 TTGCTTCATTGACTAGACN.....GACNTNAGC 2250
637 alaAspAlaGlyValAspThrGlnTrpGlyAlaLeuMetAlaAspSerSer 653
2251 GGCANTTNAGNCNTNCCNATNAGNTNNTTNAANCTCNCNGCGNTGC 2300
654 GlyGln...HisGlnAspGluGlySerThrLeuThrLysThrIleAlaG 669
2301 NNCATNAAAGCAATCTGTGCAAAATGGCGATACAGTTATACATGCA 2350
669 yThrLeu.....GluLeuThrAlaSerGlyThrThrGlnSerAlaVal. 683
2351 GCCACAAAGCCACCACCAAAACGGACCTTAGCTCGTGGCAATGCCCAA 2400
684 .....ArgValGluGluGlyThrLeuGln.....GlyAspValAla 695
2401 GCAAATTTAATCAAGCACATTA.....AAGGCGCAACATC 2438
696 AspIlePheProTyrAlaSerSerLeuTyrPvalGlyAspGlyAlaThrPh 712
2439 GGNWTGCGCAATGCTTCATTTAATCTAAGCAACACCGCGCAACAAAG 2488
712 eValThrGlyAlaAspGlnAspIleGlnSerIleAspAlaThrSerSerG 729
2489 GCACTGACGCTTTCGACACACGCTAAGCAACGTAAGCAATGCAATTC 2535
729 LysThrIleAspIleSerAspGlyThrValLeuArgLeuThrGlyGlnAsp 745
2536 .....GCACTCAACGCAATGTCTCCCTGACCGCATTAAGCAATATT 2576
746 ThrSerValAlaLeuAsnAlaSerLeuPheAsnCysAspGlyThrLeuVa 762
2577 CCAT...TTTGAACACAGCGCTTACCGGCAACTCAGCGGACGCAAG 2623
762 IAsnAlaIleThrAspGlyValThrLeuThrGlyGluLeuAsnThrAsnLeuG 779
2624 ANACAGCATTACACTTAAAGACAGCGGAATGACGCTGCGTCAGGACAG 2673

```

779 luthr.....AspSerLeuThrTyrLeuSerAsnValThr 790
2674 GAATTAAGCAATTTAACTTACCTTGACAACGGCCACCATTTACACCAATTCGGC 2723
|||||
791 ValAsnGlyAsnLeuThrAsnThrSerGlyAlaValSerLeu..... 804
2724 CTATCGCCAGATGCTGACAGCGCGCAAAACCGCAGAGTGTGACAGAGC 2773
|||||
805GlnAsnGlyAlaGlyAsp.... 811
2774 CGCGCGCGCTGCGCGCTTCCCTATTATCCGTTACACCGCCAACTTCG 2823
811 811
2824 GTAGATCCCGTTTCAACAGCGTGAACGGCAACCAATTAACAGTCA 2873
|||||
812ThrLeuThrValAsnGlyAspTyrThrGlyG1 822
2874 AGAACATTCGCTTTATGTCGGAACCTTTCGGC.....TACCGAAGC 2917
|||||
822 yGlyThrLeuLeuLeuAspSerGlnLeuAsnGlyAspAspSerValSerA 839
2918 ACAATATGAAGCTGGCGGAACTCCGGAAGNACTTACACCTTGGCGGTC 2967
|||||
839 spGlnLeuValMetAsnGlyAsnThrAlaGly**ThrThrValValAl 855
2968 AACCAAT..ACCGGC...AAGAACCCGTAAGCCTCGATCATTCACGCT 3011
|||||
856 AsnSerIleThrGlyIleGlyLeuProThrSerThrGlyIleLysValVa 872
3012 AGTGAAGAGAAAGACAAACAAACCGCTTCGAAACCTTAATTCACCC 3061
|||||
872 lAspPheAlaAlaAspProThrGlnPheGlnAsnAlaGlnPheSerL 889
3062 TGCAAAGCAA...CACGTCGATGCCGGCGCTGGCTTACCACTCATC 3108
|||||
889 euAlaGlySerGlyValAlaMetGlyAlaTyrAspTyrThrLeuVal 905
3109 CGCAAGAGCGGAGTCCGCTGCATATCCGTCAAAGAACAGAGCT 3158
|||
906 GlnAspAsnAsnAspTyrTyrLeu..... 913
3159 TTCGCAACAACCTGGCAAGGCAAGCAAAACAGCGGGAAGACA 3208
913 913
3209 AGCGGCAAAAGCCTTGACGGCTGATTCGGCGCGCGCATGCCCGCAA 3258
913 913
3259 AAGACAGAAAGCGTTGCCGACCGCGCGCGCGAGCGGGAATGT 3308
|||||
914 ArgSerGlnGlnValThrProProSer..... 922
3309 CGGCATTATGACGGCGAGAGAGAAAAACGGGTGACAGCGGATTAAG 3358
|||
923ProProAspProA 927
3359 ACAGCGCNTTGGGGAACAGCGCGAAGCGGAACCGCGCGNTACACAC 3408
|||
927 spProThr..... 929
3409 GCCTTCCCCCGCGCGCGCGCGCGGATTTGCCGCAACGCGAGCC 3458
|||||
930ProAspProAspT 934
3459 CCAACCGCAACTCAACCCCAACCGAGCGACCTGATNAGCGTTATG 3508
|||||
934 ofThrGlnAspProAspProThrProAspProGlnProThrProAlaTyrG 951
3509 CCAATAGCGGTTGAGTGAATTTCCGCGACGCTCAACAGCGTTTCGCG 3558
|||||
951 lnpValValLeuAsnAlaLysValGlyGlyTyrLeuAsnLeuThrGla 967
3559 GTACAGGACGAAATGACCGGCTGTTCGCAAGACCGCGCAAC..... 3603
|||
968 Ala.....AsnGlnAlaPheMetMetGlnTyrTrpAspHisAl 980
3604GCGTTTGGACAAGCNGCATCCGGAGACCAAC 3637
980 aGlyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyGlyAspTyrH 997
3638 ACTACCGTTGCAAGATTTCCGGCGCTACCGCCCAACAAACGACCTCGC 3687
|||||
997 lstrYThrAlaAlaGlyGlnLeuAla.....GlnHisGlyAspThrSer 1011
3688 CAAATCGTATGCGAAGAAACCTGCGCAGCGCGGC..... 3723
|||
1012 ThrValGlnLeuSerGlyAspLeuPheSerGlyArgTyrTrpLysAspG1 1028
3724GTCCGATCCCTGTTTTCGCAACCGGACCGAAGAACACT 3763
|||
1028 yGluTrpMetLeuGlyIleVal.....GlyGlyT 1038
3764 TCGACGACGGCATCGGCACTCGGACAGCGCTTGCACCGCGCGCTTTC 3813
|||||
1038 ySerAspAsnGlnGlyAspSerArgSerMetThrGlyThrArgAla 1054
3814 GGGCAATACGGCATCGGAGGTTGACATCGGCATCGACGACGAGCGCGG 3863
|||
1055 AspAsnGlnAsnHisGly...TyrAlaValGlyLeu...ThrSerTr 1069
3864 TTTTACGACGGCGACANTCTTCAGACGGCATCGCA..... 3897
|||
1069 pPheGlnHisGlyLysGlnGlyGlnGlyAlaTrpLeuAspAsnTrpLeuG 1086
3898GGCAAAATCCGCGCGCGCTG 3921
1086 lntYrAlaTrpPheSerAsnAspValSerGlnHisGlnAspGlyValAla 1102
3922 CATTAC.....GGCATTCAGCAGATACCGCGCGGTTTC..... 3957
|||||
1103 HisTyrHisSerSerGlyIleIleAlaSerLeuGlnAlaGlyTyrGlnTr 1119
3958GGCGAATTCGGCATCGAACGTCATCGCGCAACGCGCT 3997
|||
1119 pLeuProGlyArgGlyValValIleGlnProGlnAlaGlnValIleTyrG 1136
3998 ATTTCGTCAAAAGCGGATTACCGCTACGAAAC..... 4032
1136 lngLysAlGlnGlnAspPheThrAlaAlaAsnArgAlaArgValSer 1152
4033GTCAATATCGCACCCCGGCTTGCCTTCAACGNTA 4070
1153 GlnSerGlnGlyAspAspIleGlnThr..... 1161
4071 CCGNCGCGGCAATTAAAGCAGATTATTCATTCAAACCGCGCAACACATNT 4120
|||
1162 ArgLeuGlyLeuHisSerGlnTrp.....ArgThrAlaValHis.... 1174
4121 CCATCAGNCCTTATTNAGCTGTCC...TATACCGATGCCGCTTCGGGC 4167
|||
1175 ..ValIleProThrLeuAspLeuAsnTyrTyrHisAspProHisSerThr 1190
4168 AAGTCGCAACAGCGTCATATACCGCNGTATGCTCAGGATTTTCGGCAA 4217
|||||
1191 GlnIleGlnGlnAspAlaSerThr.....IleSerAspAlaValVal 1205
4218 AACCGC...AGTCCGAATGCGGCGTAAACGCCGAATC...AAAGTT 4261
|||
1205 sGlnArgGlyGlnIleLysValGlyValThrGlyAsnHisSerGlnArgV 1222
4262 TCAGCTGTCCNTCCAGCTGCCCGCGCAAGAGNCCGCAACTGGAAGCG 4311
|||||
1222 alSerLeuArgGlySerValAlaTrpGlnLysGlySerAspAspPheAla 1238

Mon Jul 1 09:26:51 2002

us-09-303-518d-651.rpr

Page 104

4312 CAACACAGCGCGGGCATCAATTAGGCTACCCTGG 4347
|| : : : : : : : : : : : : : : : :
1239 GlnThrAlaGlyPheLeuSerMetThrValLysTrp 1250